

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-060165

(43)Date of publication of application : 06.03.2001

(51)Int.Cl. G06F 12/00
G06F 13/00
G06F 17/30

(21)Application number : 11-234403

(71)Applicant : NEC CORP

(22)Date of filing : 20.08.1999

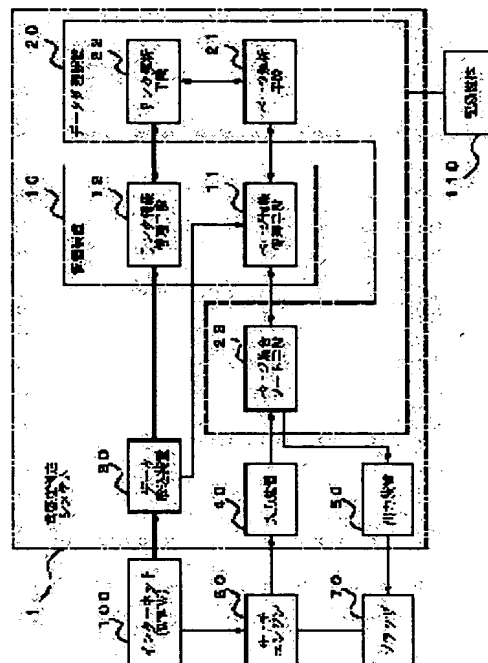
(72)Inventor : TAKANO HAJIME

(54) SYSTEM AND METHOD FOR DECIDING IMPORTANCE DEGREE OF INFORMATION SET
AND RECORDING MEDIUM RECORDING INFORMATION SET IMPORTANCE DEGREE
DISCRIMINATION PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To enable the optimal relocation of arbitrary page sets in hyper media information while using information sets edited with the same theme in an information set importance degree discrimination system objective to the hyper media information on a network.

SOLUTION: This information set importance degree DECISION system is provided with a page information managing means 11 for managing the attributes of page information composing of the hyper media information, a link information managing means 12 for holding and managing the link relation of page information, a data fetching device 30 for registering the hyper media information into the page information managing means 11 and the link information managing means 12, a page analyzing means 21 for deciding the degree of importance for each page, a link analyzing means 22 for deciding the weight of each link on the basis of the link information and a page set sort means 23 for sorting the page sets on the basis of the degree of importance of page for each page.



LEGAL STATUS

[Date of request for examination]

10.07.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-60165

(P2001-60165A)

(43) 公開日 平成13年3月6日(2001.3.6)

(51) Int.Cl.⁷

G 0 6 F 12/00

13/00

17/30

識別記号

5 4 6

3 5 4

F I

G 0 6 F 12/00

13/00

15/40

15/403

15/419

テーマコード*(参考)

5 4 6 B 5 B 0 7 5

3 5 4 D 5 B 0 8 2

3 1 0 F 5 B 0 8 9

3 4 0 B

3 2 0

審査請求 有 請求項の数32 O L (全 28 頁)

(21) 出願番号

特願平11-234403

(22) 出願日

平成11年8月20日(1999.8.20)

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 高野 元

東京都港区芝五丁目7番1号 日本電気株式会社内

(74) 代理人 100093595

弁理士 松本 正夫

Fターム(参考) 5B075 ND36 NR15 PQ02 PQ75 PR06

QM08

5B082 EA00

5B089 GA11 GA21 GB04 HA10 JA24

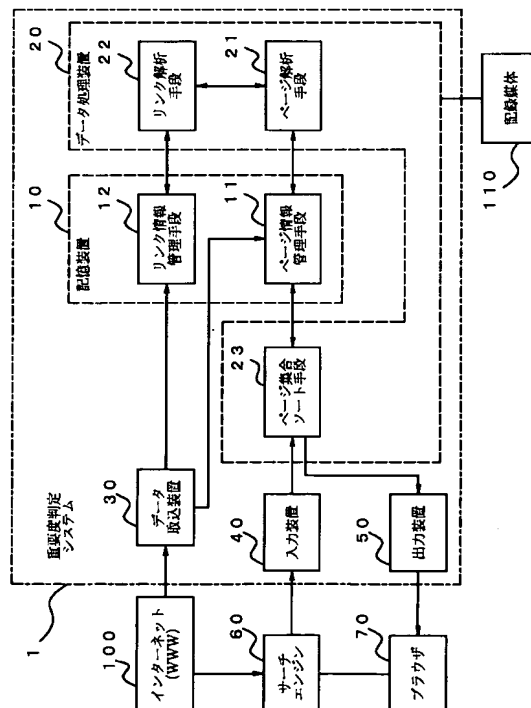
JB22 KA02 KB07 KC23 KC39

(54) 【発明の名称】 情報セット重要度判定システム及びその判定方法、及び情報セット重要度判定プログラムを記録した記録媒体

(57) 【要約】

【課題】 ネットワーク上のハイパーメディア情報に対象とした情報セット重要度判定システムにおいて、同一テーマで編集された情報セットを用いて、ハイパーメディア情報内の任意のページ集合を最適に並べ替えることを可能とする。

【解決手段】 本発明の情報セット重要度判定システムは、ハイパーメディア情報を構成するページ情報の属性を管理するページ情報管理手段11と、ページ情報間のリンク関係を保持管理するリンク情報管理手段12と、ハイパーメディア情報をページ情報管理手段11とリンク情報管理手段12に対して登録するデータ取込装置30と、各ページに対する重要度を判定するページ解析手段21と、リンク情報を元に各リンクの重みを判定するリンク解析手段22と、ページ集合を各ページに対するページ重要度に基づいてソートするページ集合ソート手段23を備える。



【特許請求の範囲】

【請求項1】 ネットワーク上のハイパーメディア情報から、各サーバ上に存在するページを収集し、前記各サーバ間にまたがる前記各ページのリンク参照状況を取得し、前記リンク参照状況を用いて各前記ページの重要度を判定し、複数の前記ページからなるページ集合を、各前記ページに対する前記ページ重要度に基いてソートすることを特徴とする情報セット重要度判定システム。

【請求項2】 前記ハイパーメディア情報を構成するページ情報の属性を管理するページ情報管理手段と、前記ページ情報間のリンク関係を保持管理するリンク情報管理手段と、前記ハイパーメディア情報を、前記ページ情報管理手段と前記リンク情報管理手段に対して登録するページ情報登録手段と、各前記ページに対する重要度を判定するページ解析手段と、前記リンク情報を元に各リンクの重みを判定するリンク解析手段と、前記ページ集合を、各前記ページに対する前記ページ重要度に基いてソートするページ集合ソート手段を備えることを特徴とする請求項1に記載の情報セット重要度判定システム。

【請求項3】 各前記ページが属する情報セットの範囲と情報セットの要素である前記ページを管理する情報セット管理手段と、各前記ページが属する情報セットを判定する情報セット判定手段を備え、前記ページ集合ソート手段は、入力装置から与えられた前記ページ集合を前記情報セットと各ページの前記重要度に基いてソートを行い、このソート結果を出力装置に出力することを特徴とする請求項1または請求項2に記載の情報セット重要度判定システム。

【請求項4】 情報セットの代表ページを判定する代表ページ判定手段を備え、前記ページ集合ソート手段は、入力装置から与えられた前記ページ集合に対して、各前記ページが属する情報セットの前記代表ページが含まれていない場合にはこれを挿入し、前記情報セットと各ページの重要度に基いてソートを行い、このソート結果を出力装置に出力することを特徴とする請求項1乃至請求項3に記載の情報セット重要度判定システム。

【請求項5】 前記代表ページ判定手段は、前記情報セット内の各ページに対し、前記情報セット外のページからのリンク数を調べ、この前記リンク数が最大のページをその情報セットの代表ページと判定することを特徴とする請求項4の情報セット重要度判定システ

ム。

【請求項6】 前記代表ページ判定手段は、前記情報セット外のページからの前記リンク数が最大であるページが複数存在する場合に、前記情報セット内のページからの前記リンク数が最大であるものをその情報セットの代表ページと判定することを特徴とする請求項5の情報セット重要度判定システム。

【請求項7】 前記代表ページの判定に用いるファイル名を格納する代表ページ候補格納手段を備え、

10 前記代表ページ判定手段は、前記情報セット外のページからの前記リンク数が最大でありかつ前記情報セット内のページからの前記リンク数が最大である前記ページが複数存在する場合に、前記代表ページ候補格納手段に格納されている前記ファイル名と前記ページの持つファイル名を順次比較し、前記ファイル名が合致したものをその情報セットの代表ページと判定することを特徴とする請求項6の情報セット重要度判定システム。

【請求項8】 同じ深さに位置する下位ディレクトリを取り出し、各前記下位ディレクトリ毎にリンク参照数が最大のページを選択し、前記リンク参照数の合計値が前記代表ページとして判定されたページのリンク参照数よりも大きい場合には、各前記下位ディレクトリに対し前記代表ページ判定処理を再帰的に実行することを特徴とする請求項5乃至請求項7に記載の情報セット重要度判定システム。

【請求項9】 前記情報セット判定手段は、前記ページ解析手段を介して前記ページ情報管理手段に格納されているページをサーバ単位で取出し、30 前記サーバ上の前記ページから前記代表ページを選択し、前記代表ページを起点として、前記代表ページと同一ディレクトリに存在するページへのリンクを順次たどり、得られた前記ページを情報セットのメンバと判定し、得られた前記情報セットを前記情報セット管理手段に登録することを特徴とする請求項4乃至請求項7に記載の情報セット重要度判定システム。

【請求項10】 前記情報セット判定手段は、前記情報セットの範囲を決定している同一サーバ上の同一ディレクトリ以下に存在し、かつ前記代表ページを起点として順次リンクを経由して到達することのできないページを前記情報セットから削除し、前記削除したページを、前記削除したページへのリンクを保持するページの属する情報セットに属するものと設定することを特徴とする請求項9の情報セット重要度判定システム。

【請求項11】 前記情報セット中の各ページの主要な関連を判定し、各前記ページの親ページと前記情報セットにおける役割を、前記ページ管理手段に設定する情報

木判定手段を備えることを特徴とする請求項6乃至請求項10に記載の情報セット重要度判定システム。

【請求項12】 前記情報木判定手段は、前記情報セット管理手段から順次情報セットとその要素であるページを取出し、取出した全前記ページについて前記情報セット内のページからのリンク数を数え、前記被リンク数が1のページは、そのリンク元ページを親ページとして設定し、次に下層から順にサブディレクトリを取出し、前記ページのサブディレクトリに属する子ページをリンクしているリンク元ページを調べ、各前記リンク元ページに対してディレクトリ内ページへのリンク数である親度数を数え、前記リンク元ページのうち最大のリンク数を持つものから順に、この前記リンク元ページにリンクされている各子ページの親ページとして設定し、全てのページに上記一連の処理を行い親ページを設定することを特徴とする請求項11に記載の情報セット重要度判定システム。

【請求項13】 前記情報木判定手段は、前記情報セット内の全ページに対して、各前記ページを親ページとして設定している子ページの数である親度数を数え、前記親度数が予め決められた値よりも大きなページを”インデックス”とラベル付けすることを特徴とする請求項12に記載の情報セット重要度判定システム。

【請求項14】 関連の強い前記情報セットを選別する情報セット選択手段と、前記情報セット選択手段が選択した複数の前記情報セットを集約して新たな情報セットを生成する情報セット集約手段を備えることを特徴とする請求項1乃至請求項14に記載の情報セット重要度判定システム。

【請求項15】 前記情報セット選択手段は、前記情報セットの属するネットワーク上の位置の重なりに基づいて新たな情報セットを生成することを特徴とする請求項14に記載の情報セット重要度判定システム。

【請求項16】 前記情報セット選択手段は、前記情報セット間に存在するリンク数を数え、前記リンク数が一定数を越えたものを新たな情報セットとすることを特徴とする請求項14に記載の情報セット重要度判定システム。

【請求項17】 前記情報セット選択手段は、前記情報セットの属するネットワーク上の位置の重なりと、前記情報セット間に存在するリンク数を用いて複数の前記情報セット間の関連強度を計算し、前記関連強度が一定数以上の前記情報セットの集合を新たな情報セットとすることを特徴とする請求項14に記載の情報セット重要度判定システム。

【請求項18】 前記情報セットとその前記情報木の構

造にしたがい、与えられた前記ページ集合をソートして出力するページ集合ソート手段を備えることを特徴とする請求項11乃至請求項13に記載の情報セット重要度判定システム。

【請求項19】 前記ページ集合ソート手段は、与えられた前記ページ集合を前記情報セット単位にグループを作成し、各前記グループに属するページに予め設定されている前記重要度を調べ、

10 その最大値をグループ重要度として設定し、各前記グループを前記グループ重要度に基づいてソートし、この前記グループのソート結果に基づいて順次前記グループ内のページをその前記ページ重要度順で出力することを特徴とする請求項18に記載の情報セット重要度判定システム。

【請求項20】 前記ページ集合ソート手段は、各前記グループに属するページに予め設定されている前記重要度を調べ、

20 その最大値をグループ重要度として設定し、各前記グループを前記グループ重要度に基づいてソートし、かつ、前記グループに所属している各ページの親ページを調べ、

前記親ページが、代表ページかまたは”インデックス”ラベルを持つページであり、かつグループに存在しない場合にはこれをグループに挿入し、前記グループのソート結果に基づいて順次前記グループ内のページをその前記ページ重要度順で出力することを特徴とする請求項18に記載の情報セット重要度判定システム。

【請求項21】 前記ページ集合ソート手段は、前記グループ内で前記”インデックス”ラベルを持つページからリンクされている子ページの数がある一定値よりも多い場合に、前記子ページの内部前記ページ重要度が上位の一定数以外を削除して出力することを特徴とする請求項20に記載の情報セット重要度判定システム。

【請求項22】 ネットワーク上のハイパーメディア情報を構成するページ情報の属性をページ情報管理手段に、前記ページ情報間のリンク関係をリンク情報管理手段に格納し、

40 各前記ページに相互のリンク参照を調べることで各リンクの重みを判定し、得られた前記リンクの重みと前記ページ情報を用いて各前記ページの重要度を判定し、入力装置から与えられた複数のページからなるページ集合を、各前記ページの前記重要度に基づいてソートし、ソート結果を出力装置に出力することを特徴とする情報セット重要度判定方法。

【請求項23】 前記ページ情報管理手段に格納されている前記ページ情報をサーバ単位で取出し、前記サーバ上のルートディレクトリ以下を前記サーバ上の第1の情報セットと判定し、前記ルートディレクトリ直下のページについて前記サーバ外のページからのリンク数を調べ、この被リンク数が最大のページを前記第1の情報セットの代表ページと判定し、前記入力装置から与えられた前記ページ集合に対して、各前記ページが属する情報セットの前記代表ページが含まれていない場合にはこれを挿入し、前記情報セットと各前記ページの前記重要度に基づいてソートし、ソート結果を出力装置に出力することを特徴とする請求項22に記載の情報セット重要度判定方法。

【請求項24】 前記情報セット管理手段から順次情報セットとその要素となるページを取出し、取出した全ページについて情報セット内のページからのリンク数を数え、前記被リンク数が1のページは、そのリンク元ページを親ページとして設定し、次に下層から順にサブディレクトリを取出し、前記ページのサブディレクトリに属する子ページをリンクしているリンク元ページを調べ、各前記リンク元ページに対してディレクトリ内ページへのリンク数である親度数を数え、前記リンク元ページのうち最大のリンク数を持つものから順に、この前記リンク元ページにリンクされている各子ページの親ページとして設定し、全てのページに上記一連の処理を行い親ページを設定することを特徴とする請求項22に記載の情報セット重要度判定方法。

【請求項25】 与えられたページ集合を情報セット単位にグループを作成し、各グループに属するページに予め設定されている重要度を調べ、その最大値をグループ重要度に設定し、各グループを前記グループ重要度に基づいてソートし、このグループソート結果に基づいて順次グループ内のページをそのページ重要度順で出力することを特徴とする請求項22に記載の情報セット重要度判定方法。

【請求項26】 各前記グループに属するページに予め設定されている前記重要度を調べ、その最大値を前記グループ重要度に設定し、これらのグループを前記グループ重要度に基づいてソートし、グループに所属している各ページの親ページを調べ、前述の親ページが代表ページであるかまたは”インデックス”ラベルを持ちグループに存在しない場合にはこれをグループに挿入し、グループソート結果に基づいて順次グループ内のページを

そのページ重要度順で出力し、前記グループ内で”インデックス”ラベルを持つページからリンクされている子ページの数値が一定値よりも多い場合には、前記子ページの内前記ページ重要度が上位の一定数以外は削除して出力することを特徴とする請求項25に記載の情報セット重要度判定方法。

【請求項27】 ネットワーク上のハイパーメディア情報に関し、各サーバ上に存在するページを収集し、各前記サーバの内側及び外側にまたがる前記各ページのリンク参照状況を調べる処理と、前記リンク参照状況を用いて各前記ページの重要度を判定する処理と、複数の前記ページからなるページ集合を、各前記ページに対する前記ページ重要度に基づいてソートする処理を、コンピュータに実行させるための情報セット重要度判定プログラムを記録したことを特徴とする記録媒体。

【請求項28】 前記ハイパーメディア情報を構成するページ情報の属性を管理するページ情報管理処理と、前記ページ情報間のリンク関係を保持管理するリンク情報管理処理と、前記ハイパーメディア情報を前記ページ情報管理処理と前記リンク情報管理処理に対し登録するページ情報登録処理と、各前記ページに対する重要度を判定するページ解析処理と、前記リンク情報を元に各リンクの重みを判定するリンク解析処理と、前記ページ集合を、各前記ページに対する前記ページ重要度に基づいてソートするページ集合ソート処理をコンピュータに実行させるための情報セット重要度判定プログラムを記録したことを特徴とする記録媒体。

【請求項29】 各前記ページが属する情報セットの範囲と情報セットの要素である前記ページを管理する情報セット管理処理と、各前記ページが属する情報セットを判定する情報セット判定処理を備え、前記ページ集合ソート処理は、入力装置から与えられた前記ページ集合を前記情報セットと各ページの前記重要度に基づいてソートを行い、このソート結果を出力装置に出力することを特徴とする請求項27または請求項28に記載の記録媒体。

【請求項30】 前記情報セット中の各ページの主要な関連を判定し、各前記ページの親ページと前記情報セットにおける役割を、前記ページ管理処理に設定する情報木判定処理を備え、前記情報木判定処理は、前記情報セット管理処理から順次情報セットとその要素であるページを取出し、取出した全前記ページについて前記情報セット内のペー

ジからのリンク数を数え、
前記被リンク数が1のページは、そのリンク元ページを親ページとして設定し、
次に下層から順にサブディレクトリを取出し、
前記ページのサブディレクトリに属する子ページをリンクしているリンク元ページを調べ、
各前記リンク元ページに対してディレクトリ内ページへのリンク数である親度数を数え、
前記リンク元ページのうち最大のリンク数を持つものから順に、この前記リンク元ページにリンクされている各子ページの親ページとして設定し、
全てのページに上記一連の処理を行い親ページを設定することを特徴とする請求項29に記載の記録媒体。

【請求項31】 前記情報セットとその前記情報木の構造にしたがい、与えられた前記ページ集合をソートして出力するページ集合ソート処理を備え、
前記ページ集合ソート処理は、
与えられた前記ページ集合を前記情報セット単位にグループを作成し、
各前記グループに属するページに予め設定されている前記重要度を調べ、
その最大値をグループ重要度として設定し、
各前記グループを前記グループ重要度に基づいてソートし、
この前記グループのソート結果に基づいて順次前記グループ内のページをその前記ページ重要度順で出力することを特徴とする請求項29に記載の記録媒体。
【請求項32】 前記ページ集合ソート処理は、
各前記グループに属するページに予め設定されている前記重要度を調べ、
その最大値をグループ重要度として設定し、
各前記グループを前記グループ重要度に基づいてソートし、
かつ、前記グループに所属している各ページの親ページを調べ、
前記親ページが、代表ページかまたは”インデックス”ラベルを持つページであり、かつグループに存在しない場合にはこれをグループに挿入し、
前記グループのソート結果に基づいて順次前記グループ内のページをその前記ページ重要度順で出力し、
前記ページ集合ソート処理は、
前記グループ内で前記”インデックス”ラベルを持つページからリンクされている子ページの数がある一定値よりも多い場合に、
前記子ページの内部前記ページ重要度が上位の一定数以外を削除して出力することを特徴とする請求項31に記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はネットワーク上のハ

イパーメディア情報における情報セット重要度判定システム、情報セット重要度判定方法及び情報セットとその重要度判定用プログラムを記録した記録媒体に関し、特に与えられたページリストについて、利用者が必要とするページ情報を上位に提示することで情報の選別に要する手間を削減することを可能とした情報セット重要度判定システム、情報セット重要度判定方法及び情報セットとその重要度判定用プログラムを記録した記録媒体に関する。

【0002】

【従来の技術】近年、コンピュータネットワークを用いた情報発信・共有の仕組みが著しく普及し、小規模なLAN（ローカルエリアネットワーク）環境から中規模なイントラネット環境、さらには日本全体あるいは世界全体を網羅する規模での利用が可能となっている。これらの情報は、情報の断片であるページと、その関係であるリンクからなるハイパーメディア情報を構成し、任意のページからいくつもの関連情報のページを逐次的にたどっていくことができる。

【0003】上記のハイパーメディア情報システムの代表的なものはWWW（ワールドワイド・ウェブ）である。その構成要素は、ページの記述フォーマットを規定するHTML（ハイパーテキスト・マークアップ・ランゲージ）、ページのネットワーク上のアドレスを規定するURL（ユニバーサル・リソース・ロケータ）、及びURLが指すアドレスに存在するページをダウンロードするための通信手順であるHTTP（ハイパーテキスト・トランスファー・プロトコル）の3つである。また、情報ページをダウンロードして閲覧するシステムをWWWブラウザ（以下ブラウザ）、情報ページを管理しブラウザの要求に対してHTTPを介してページを転送するシステムをWWWサーバ（以下サーバ）と呼んでいる。WWWはすでに、世界中に張り巡らされているネットワークであるインターネットにおいて必須のアプリケーションとなっており、全世界で数億ページが存在しているといわれている。

【0004】WWWのように極めて広範囲かつ大量に存在するページの中から、所望の情報を引き出すことは極めて難しく、そのために様々なサーチエンジンが開発され広く利用されている。サーチエンジンは、ネットワーク上のページのHTMLを解析し、そこから張られているリンク先のURLを得るという一連の作業を繰返すことでネットワーク上に存在するページを収集するクローラと、収集したページに基づいてたとえばキーワードによる全文検索を実施する検索エンジンとで構成される。

【0005】このため、たとえば日本国内の数百万以上にもなるページを収集し、これに対してキーワード検索を実施したとすると、そのほとんどの検索結果は数百から数万件というオーダになってしまう。さらにこうして出力される検索結果は、そのアドレスや内容に基づいて整

理されていないため、検索結果のブラウジングはユーザにとって極めて苦痛な作業となってしまう。

【0006】ブラウジング作業を効率化するためには、ユーザに出力すべき情報の構造を適切に判別して、情報を整理、選別、加工する必要がある。すなわち、適切な情報セットの範囲を設定できること、情報セットの代表ページと主要な構造を発見すること、情報セット内の各情報を取捨選択・加工すること、ができなければならない。ここで、情報セットとは、検索等により選別された特定のページ集合に関する情報（ページ情報、リンク情報）のセットであり、代表ページとそのメンバのページから構成される。

【0007】これまでも、このような観点で従来技術がいくつか提案されている。以下のその従来例を示す。

【0008】第1の従来技術の例が、特開平10-069423号公報に記載されている。この従来の情報セット重要度判定システムは、ネットワーク上に存在するハイパーメディア情報のページ属性値を集中管理するディレクトリサーバを、ハイパーメディア情報のページ属性値を管理するディレクトリ情報格納部と、この属性値から二次情報を生成する二次情報生成部と、二次情報で生成した機能を備えている。この従来の情報セット重要度判定システムにおいては、ホスト名やディレクトリ名を用いてハイパーメディアの情報セットを生成している。また、得られた情報セットを、ホスト名やドメイン名などのネットワークの構成を用いて集約している。ただし、この技術はページの所在だけを考慮して情報セットを判定しているために、サーバごとに異なるページの管理ポリシーによっては適切な判定をすることができない。また、ネットワークの構成を用いて集約すべき情報セットを判定しているために、同一ドメイン内に多数の情報セットがあった場合に適切な判定をすることができない。

【0009】第2の従来技術の例が、1993年6月のプロシーディングス・オブ・16thアニュアル・インターナショナル・エーシーエム・シグアイアール・カンファレンスpp. 116-125 (Proceedings of the sixteenth annual international ACM SIGIR conference) に記載されている。この従来

の情報セット重要度判定方式においては、ページ間のリンク関係、特にその連結の強さを用いて情報セットの範囲を判定している。ただし、この技術は同一の編集者が編集したハイパーメディア情報への適用だけを考慮しており、かつ情報ページ間の関連だけを考慮して情報セットを判定しているため、WWWのようなネットワーク上のハイパーメディア情報への適用は難しい。

【0010】第3の従来技術の例が、1998年6月のプロシーディングス・オブ・9thエーシーエムカンファレンス・オン・ハイパーテキスト・アンド・ハイパー

メディア、pp. 297-298 (ACM Hypertext 98) に記載されている。この従来の情報セット重要度判定方法においては、ホスト名やディレクトリ名を用いて情報セットを生成するだけでなく、予め用意したファイル名への合致によってその代表ページを判定している。ただし、この方法では予め登録しておいたファイル名以外の情報ページを代表ページと判断することができない。

【0011】第4の従来技術の例が、特開平10-105550号公報に記載されている。この従来技術においては、ハイパーメディア情報を作成する際に、各ページ間の階層構造を規定することによって、情報ページをブラウジングする際に情報ページ間の階層構造を意識した利用ができるようになっている。ただし、この方法では予め本技術を用いて編集したハイパーメディア情報における情報構造しか利用できないことになり、既存の他の方法で編集された情報ページ間の構造を利用することはできない。

【0012】第5の従来技術の例として、近年実用化されているWWWサーチエンジンの例としてエキサイト・ジャパン (<http://www.excite.co.jp/>) を挙げるができる。この従来技術においては、キーワード検索の結果を出力する際に、ホスト名でグループ化することによって、利用者が検索結果をブラウジングする際の便宜を図っている。ただし、この方法では情報セットとホストが一致するとは限らないし、検索結果に含まれていなければホームページが提示されることはない。また、メーリングリストのメッセージなどほとんど似た内容のページが大量に検索結果に含まれる場合には、これがそのまま出力されてしまう。さらに、本来利用者のブラウジングを助けるそのインデックスページは、検索結果に含まれていなければ出力されない。

【0013】

【発明が解決しようとする課題】従来技術の第1の例と第2の例に示したように、第1の問題点は、情報セットの範囲の判定において情報作成者が意図した範囲を正確に判定できないということである。

【0014】その理由は、ホスト名やディレクトリ名のみ、あるいはリンク関係のみという限定された属性情報のみを使用しているためである。

【0015】従来技術の第3の例と第4の例に示したように、第2の問題点は、情報作成者が対象とした情報セットの代表ページを判定できないということである。

【0016】その理由は、各ページにはそのような情報構造を示す属性値が付与されていないためである。

【0017】従来技術の第5の例に示したように、第3の問題点は、与えられた情報ページの集合において、ページを取捨選択する基準がないということである。

【0018】その理由は、情報セットにおける情報構造

が判定されていないためである。

【0019】本発明の第1の目的は、ネットワーク上のハイパーメディア情報を対象とした情報セット重要度判定システムにおいて、与えられたページリストを利用者にとって価値の高いものの順にソートし、それにより利用者が必要とするページ情報を上位に提示することで情報の選別に要する手間を削減することができる情報セット重要度判定システム、情報セット重要度判定方法及び情報セットとその重要度判定用プログラムを記録した記録媒体を提供することにある。

【0020】本発明の第2の目的は、ネットワーク上のハイパーメディア情報を対象とした情報セット重要度判定システムにおいて、与えられたページリストに適切な代表ページが含まれていない場合に、代表ページを挿入してソートすることにより、検索時に適しなかった代表ページを利用者に提示することができる情報セット重要度判定システム、情報セット重要度判定方法及び情報セットとその重要度判定用プログラムを記録した記録媒体を提供することにある。

【0021】本発明の第3の目的は、ネットワーク上のハイパーメディア情報を対象とした情報セット重要度判定システムにおいて、与えられたページリストから不要なページを削除することにより、利用者に提示する検索結果の数を削減することができる情報セット重要度判定システム、情報セット重要度判定方法及び情報セットとその重要度判定用プログラムを記録した記録媒体を提供することにある。

【0022】

【課題を解決するための手段】上記目的を達成するため本発明の情報セット重要度判定システムは、ネットワーク上のハイパーメディア情報から、各サーバ上に存在するページを収集し、前記各サーバ間にまたがる前記各ページのリンク参照状況を取得し、前記リンク参照状況を用いて各前記ページの重要度を判定し、複数の前記ページからなるページ集合を、各前記ページに対する前記ページ重要度に基づいてソートすることを特徴とする。

【0023】請求項2の本発明の情報セット重要度判定システムは、前記ハイパーメディア情報を構成するページ情報の属性を管理するページ情報管理手段と、前記ページ情報間のリンク関係を保持管理するリンク情報管理手段と、前記ハイパーメディア情報を、前記ページ情報管理手段と前記リンク情報管理手段に対して登録するページ情報登録手段と、各前記ページに対する重要度を判定するページ解析手段と、前記リンク情報を元に各リンクの重みを判定するリンク解析手段と、前記ページ集合を、各前記ページに対する前記ページ重要度に基づいてソートするページ集合ソート手段を備えることを特徴とする。

【0024】請求項3の本発明の情報セット重要度判定システムは、各前記ページが属する情報セットの範囲と

情報セットの要素である前記ページを管理する情報セット管理手段と、各前記ページが属する情報セットを判定する情報セット判定手段を備え、前記ページ集合ソート手段は、入力装置から与えられた前記ページ集合を前記情報セットと各ページの前記重要度に基づいてソートを行い、このソート結果を出力装置に出力することを特徴とする。

【0025】請求項4の本発明の情報セット重要度判定システムは、情報セットの代表ページを判定する代表ページ判定手段を備え、前記ページ集合ソート手段は、入力装置から与えられた前記ページ集合に対して、各前記ページが属する情報セットの前記代表ページが含まれていない場合にはこれを挿入し、前記情報セットと各ページの重要度に基づいてソートを行い、このソート結果を出力装置に出力することを特徴とする。

【0026】請求項5の本発明の情報セット重要度判定システムにおける前記代表ページ判定手段は、前記情報セット内の各ページに対し、前記情報セット外のページからのリンク数を調べ、この前記リンク数が最大のページをその情報セットの代表ページと判定することを特徴とする。

【0027】請求項6の本発明の情報セット重要度判定システムにおける前記代表ページ判定手段は、前記情報セット外のページからの前記リンク数が最大であるページが複数存在する場合に、前記情報セット内のページからの前記リンク数が最大であるものをその情報セットの代表ページと判定することを特徴とする。

【0028】請求項7の本発明の情報セット重要度判定システムは、前記代表ページの判定に用いるファイル名を格納する代表ページ候補格納手段を備え、前記代表ページ判定手段は、前記情報セット外のページからの前記リンク数が最大でありかつ前記情報セット内のページからの前記リンク数が最大である前記ページが複数存在する場合に、前記代表ページ候補格納手段に格納されている前記ファイル名と前記ページの持つファイル名を順次比較し、前記ファイル名が合致したものをその情報セットの代表ページと判定することを特徴とする。

【0029】請求項8の本発明の情報セット重要度判定システムは、同じ深さに位置する下位ディレクトリを取り出し、各前記下位ディレクトリ毎にリンク参照数が最大のページを選択し、前記リンク参照数の合計値が前記代表ページとして判定されたページのリンク参照数よりも大きい場合には、各前記下位ディレクトリに対し前記代表ページ判定処理を再帰的に実行することを特徴とする。

【0030】請求項9の本発明の情報セット重要度判定システムにおける前記情報セット判定手段は、前記ページ解析手段を介して前記ページ情報管理手段に格納されているページをサーバ単位で取出し、前記サーバ上の前記ページから前記代表ページを選択し、前記代表ページ

を起点として、前記代表ページと同一ディレクトリに存在するページへのリンクを順次たどり、得られた前記ページを情報セットのメンバと判定し、得られた前記情報セットを前記情報セット管理手段に登録することを特徴とする。

【0031】請求項10の本発明の情報セット重要度判定システムにおける前記情報セット判定手段は、前記情報セットの範囲を決定している同一サーバ上の同一ディレクトリ以下に存在し、かつ前記代表ページを起点として順次リンクを経由して到達することのできないページを前記情報セットから削除し、前記削除したページを、前記削除したページへのリンクを保持するページの属する情報セットに属するものと設定することを特徴とする。

【0032】請求項11の本発明の情報セット重要度判定システムは、前記情報セット中の各ページの主要な関連を判定し、各前記ページの親ページと前記情報セットにおける役割を、前記ページ管理手段に設定する情報木判定手段を備えることを特徴とする。

【0033】請求項12の本発明の情報セット重要度判定システムにおける前記情報木判定手段は、前記情報セット管理手段から順次情報セットとその要素であるページを取出し、取出した全前記ページについて前記情報セット内のページからのリンク数を数え、前記被リンク数が1のページは、そのリンク元ページを親ページとして設定し、次に下層から順にサブディレクトリを取出し、前記ページのサブディレクトリに属する子ページをリンクしているリンク元ページを調べ、各前記リンク元ページに対してディレクトリ内ページへのリンク数である親度数を数え、前記リンク元ページのうち最大のリンク数を持つものから順に、この前記リンク元ページにリンクされている各子ページの親ページとして設定し、全てのページに上記一連の処理を行い親ページを設定することを特徴とする。

【0034】請求項13の本発明の情報セット重要度判定システムにおける前記情報木判定手段は、前記情報セット内の全ページに対して、各前記ページを親ページとして設定している子ページの数である親度数を数え、前記親度数が予め決められた値よりも大きなページを”インデックス”とラベル付けすることを特徴とする。

【0035】請求項14の本発明の情報セット重要度判定システムは、関連の強い前記情報セットを選別する情報セット選択手段と、前記情報セット選択手段が選択した複数の前記情報セットを集約して新たな情報セットを生成する情報セット集約手段を備えることを特徴とする。

【0036】請求項15の本発明の情報セット重要度判定システムにおける前記情報セット選択手段は、前記情報セットの属するネットワーク上の位置の重なりに基づいて新たな情報セットを生成することを特徴とする。

【0037】請求項16の本発明の情報セット重要度判定システムにおける前記情報セット選択手段は、前記情報セット間に存在するリンク数を数え、前記リンク数が一定数を超えたものを新たな情報セットとすることを特徴とする。

【0038】請求項17の本発明の情報セット重要度判定システムにおける前記情報セット選択手段は、前記情報セットの属するネットワーク上の位置の重なりと、前記情報セット間に存在するリンク数を用いて複数の前記情報セット間の関連強度を計算し、前記関連強度が一定数以上の前記情報セットの集合を新たな情報セットとすることを特徴とする。

【0039】請求項18の本発明の情報セット重要度判定システムは、前記情報セットとその前記情報木の構造にしたがい、与えられた前記ページ集合をソートして出力するページ集合ソート手段を備えることを特徴とする。

【0040】請求項19の本発明の情報セット重要度判定システムにおける前記ページ集合ソート手段は、与えられた前記ページ集合を前記情報セット単位にグループを作成し、各前記グループに属するページに予め設定されている前記重要度を調べ、その最大値をグループ重要度として設定し、各前記グループを前記グループ重要度に基づいてソートし、この前記グループのソート結果に基づいて順次前記グループ内のページをその前記ページ重要度順で出力することを特徴とする。

【0041】請求項20の本発明の情報セット重要度判定システムにおける前記ページ集合ソート手段は、各前記グループに属するページに予め設定されている前記重要度を調べ、その最大値をグループ重要度として設定し、各前記グループを前記グループ重要度に基づいてソートし、かつ、前記グループに所属している各ページの親ページを調べ、前記親ページが、代表ページかまたは”インデックス”ラベルを持つページであり、かつグループに存在しない場合にはこれをグループに挿入し、前記グループのソート結果に基づいて順次前記グループ内のページをその前記ページ重要度順で出力することを特徴とする。

【0042】請求項21の本発明の情報セット重要度判定システムにおける前記ページ集合ソート手段は、前記グループ内で前記”インデックス”ラベルを持つページからリンクされている子ページの数がある一定値よりも多い場合に、前記子ページの内部前記ページ重要度が上位の一定数以外を削除して出力することを特徴とする。

【0043】請求項22の本発明の情報セット重要度判定方法は、ネットワーク上のハイパーメディア情報を構成するページ情報の属性をページ情報管理手段に、前記ページ情報間のリンク関係をリンク情報管理手段に格納し、各前記ページに相互のリンク参照を調べることで各リンクの重みを判定し、得られた前記リンクの重みと前

記ページ情報を用いて各前記ページの重要度を判定し、入力装置から与えられた複数のページからなるページ集合を、各前記ページの前記重要度に基づいてソートし、ソート結果を出力装置に出力することを特徴とする。

【0044】請求項27の本発明の記録媒体は、ネットワーク上のハイパーメディア情報に関し、各サーバ上に存在するページを収集し、各前記サーバの内側及び外側にまたがる前記各ページのリンク参照状況を調べる処理と、前記リンク参照状況を用いて各前記ページの重要度を判定する処理と、複数の前記ページからなるページ集合を、各前記ページに対する前記ページ重要度に基づいてソートする処理を、コンピュータに実行させるための情報セット重要度判定プログラムを記録したことを特徴とする。

【0045】

【発明の実施の形態】以下、本発明の実施の形態について図面を参照して詳細に説明する。図1は、本発明の第1の実施の形態による重要度判定システムの構成を示すブロック図である。

【0046】図1を参照すると、本発明の第1の実施の形態による重要度判定システム1は、データを記憶する記憶装置10と、プログラム制御により動作するデータ処理装置20と、インターネット(WWW)100上のファイルやデータベースからデータを取り込んで記憶装置10に登録するデータ取込装置30と、インターネット100を利用するブラウザ70からサーチエンジン60によって検索された検索結果(URLリスト)を入力する入力装置40と、ディスプレイやファイル等の出力装置50を備える。

【0047】記憶装置10は、ページ情報管理手段11と、リンク情報管理手段12を備える。

【0048】ページ情報管理手段11は、データ取込装置30によって記憶装置10に登録された重要度判定対象となるページのページ情報を格納する。

【0049】リンク情報管理手段12は、データ取込装置30によって記憶装置10に登録された重要度判定対象であるページ間のリンク情報を格納する。

【0050】データ処理装置20は、ページ解析手段21と、リンク解析手段22と、ページ集合ソート手段23を備える。

【0051】ページ解析手段21は、ページ情報管理手段11から解析対象のページ集合を取出し、取出した個々のページ集合のページ重要度を計算し、その結果をページ情報管理手段11に対し格納する。

【0052】リンク解析手段22は、ページ解析手段21から指定されたページ情報について、リンク情報管理手段12から関連するリンク情報を取出し、取出した個々のリンクの重要度を判定し、その結果をリンク情報管理手段12に格納する。

【0053】ページ集合ソート手段23は、入力装置4

0を介して渡された、例えばサーチエンジン60によって検索された検索結果であるページ集合を、情報管理手段11に格納されている重要度に従ってソートし、ソート結果を出力装置50に対し出力する。

【0054】次に、本実施の形態の重要度判定の処理について図面を参照して詳細に説明する。図2は、本発明の第1の実施の形態の重要度判定の処理を説明するためのフローチャートである。

【0055】データ取込装置30は、予めWWWなどのネットワーク上のハイパーメディア情報を収集して作成したデータベースに格納されているページ情報とリンク情報を、それぞれページ情報管理手段11とリンク情報管理手段12に登録する。ページ情報を構成する属性の1つにネットワーク上でのアドレス、たとえばWWWにおけるURLが”http://www.nec.co.jp/”のように設定されているものとする。

【0056】ここで、ネットワーク上のページアドレスをURLと記すこととする。よって、リンク情報はリンク元URLとリンク先URLの対で1レコードを構成するものとする。

【0057】本実施の形態の重要度判定の処理においては、各変数を以下のように定める。

【0058】あるURLにおいてこのリンク先URLを(N個所ある場合には)DP(i)(i=0、…、N-1)と、リンク先URLの1つDP(i)に対するリンク元URLを(M個所ある場合には)SP(i,j)(j=0、…、M-1)と、リンク先URLの1つDP(i)の重要度をIMP(i)と表す。

【0059】ページ解析手段21は、ページ情報管理手段11から調査すべきページのURLをまとめたURLリストを得る(ステップ201)。

【0060】次にページ解析手段21はURLリストから順次URLを取出して、これをリンク解析手段22に通知する。

【0061】リンク解析手段22は、通知されたURLのリンク先URLであるDP(i)(i=0、…、N-1)をリンク情報管理手段12から取出し(ステップ202)、リンク解析手段22を通じて取出したDP(i)へのリンクを保持している被リンクURLリストを得る(ステップ203)。

【0062】リンク解析手段22は、被リンクURLリストからリンク元URLであるSP(i,j)を順次取出し(ステップ204)、その重要度を算出する(ステップ205)。

【0063】ここで、重要度の算出においては、リンク元URLであるSP(i,j)(j=0、…、M-1)とリンク先URLであるDP(i)が異なるサーバ上であれば重要度を高くし、同一サーバ上であれば重要度を低く設定する方式を適用する。具体的には、リンク元URLであるSP(i,j)(j=0、…、M-1)とり

リンク先URLであるDP (i) が異なるサーバ上にある場合、重要度に"1"を加算し、同一サーバ上にある場合には、何も値を加算しない。ここで計算された重要度は、リンク先URLのDP (i) の重要度IMP (i) に加算される(ステップ206)。

【0064】なお、ページの重要度の設定方法については、上記の方法に限定されずその他の方法を適用することができる。これまでの重要度設定方法としては、例えば、以下に示すものが知られており、本発明においてもそれらの方法を利用するものとする。

【0065】第1に、Sougata Mukherjea and Yoshinori Hara, "Focus+Content Views ofWorld-Wide Web Node s", Proceedings of ACM Hypertext'97, ACM Press, pp. 187-196には、一回又は二回のクリックで辿れる文書

(ページ)の数をカウントし、これをページの重要度として設定する方法が示されている。第2に、Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" pp.107-117, Proceedings of 7th WWWConf., May 1998には、相互接続されたリンク関係の全体構造から、個々のページ間参照の強さを算出し、それをページ重要度とする方法が示されている。

【0066】リンク解析手段22は、DP (i) に対するリンク元URLである全てのSP (i, j) に対しステップ204からステップ207の処理を実行する(ステップ207)。

【0067】処理が終わった時点で重要度IMP (i) の値をページ解析手段21に通知し、ページ解析手段21は、通知された重要度IMP (i) をページDP (i) の重要度としてページ情報管理手段11に設定する(ステップ208)。

【0068】引き続きページ解析手段21は、URLリストDP (i) (i=0、…、N-1)の残る全てに対しステップ202からステップ208までの処理を実行する(ステップ209)。

【0069】以上により、ページ情報管理手段11には各URLのページの重要度が設定される。

【0070】ページ集合ソート手段23は、入力装置40を介して入力された、例えばサーチエンジン60によって検索された検索結果であるURLの集合から、個々のURLに対応するページの重要度をページ情報管理手段11から得て、さらにページ集合をその重要度でソートし、その結果が出力装置50であるディスプレイやファイル等に出力される。

【0071】以上のように、本実施の形態では、各ページのネットワークアドレスだけでなくリンク関係に基づいた重要度を判定する手段を備えるために、単にページ自体が持つ特徴から計算したものではなく、他の関連情報との関係も含めたページの重要度を得ることができる。

【0072】なお、本実施の形態では、図2のステップ205における重要度計算を、リンク元URLとリンク先URLの属するサーバの一致度で判定した。他にも、リンク先URLが存在するサーバと、リンク元URLが存在するサーバ間にリンクが存在すれば、そのリンクの数がいくつであっても重要度に"1"を加算するという方法も考えられる。また、サーバの同一性だけではなく、ディレクトリの同一性も同時に考慮して重要度を算出する方法も考えられる。

10 【0073】次に、本発明の第2の実施の形態について図面を参照して詳細に説明する。図3は、本発明の第2の実施の形態による重要度判定システム1の構成を示すブロック図である。

【0074】図3を参照すると、本発明の第2の実施の形態による重要度判定システムの、図1に示された第1の実施の形態による重要度判定システムとの違いは、記憶装置10aにおいて、図1に示された第1の実施の形態の記憶装置10の構成に加え情報セット管理手段13を備え、データ処理装置20aにおいて、図1に示された第1の実施の形態のデータ処理装置20の構成に加え情報セット判定手段24と代表ページ判定手段25を備え、かつ第1の実施の形態のページ集合ソート手段23を拡張したページ集合ソート手段23aを備える点である。

【0075】情報セット管理手段13は、代表ページとそのメンバのページから構成される情報セットの範囲と、その構成メンバであるURLのリストを保持している。

【0076】図4は、本発明の第2の実施の形態の情報セット判定手段24の構成を示すブロック図である。

【0077】図4を参照すると、第2の実施の形態の情報セット判定手段24は、URLリスト作成手段24-1と、一時作業領域24-2と、代表ページ選択手段24-3と、バウンダリ判定手段24-4と、親子探索手段24-5と、他セット移籍手段24-6と、情報セット作成手段24-7を備える。

【0078】情報セット判定手段24は、ページ解析手段21とリンク解析手段22を用いて、ページ情報管理手段11に格納されている全ページを対象として適切な情報セットを判定する。さらに、情報セット判定手段24は、各ページの重要度を情報セット管理手段13に格納されている情報セットのバウンダリに基づいて判定する。バウンダリとは、代表ページとそのメンバのページから構成されるある情報セットに関し、メンバとして適合するための条件を示す。ここで、情報セットとバウンダリ235の概念を図23に示す。図23において、点線で示されるのがバウンダリ235であり、情報セット230の境界を示している。また、各情報セット230は、それぞれ代表ページ231と複数のメンバURL232で構成されている。図において、矢印はリンクを示してい

る。

【0079】図5は、第2の実施の形態の代表ページ判定手段25の構成を示すブロック図である。

【0080】図5を参照すると、代表ページ判定手段25は、サーバリスト保持手段25-1と、メンバURL取得手段25-2と、一時作業領域25-3と、サーバ外被リンク数カウント手段25-4と、ディレクトリ切出し手段25-5と、代表ページ選択手段25-6と、サーバ内被リンク数カウント手段25-7と、代表候補名格納手段25-8と、代表ページ登録手段25-9を備える。

【0081】代表ページ判定手段25は、ページ解析手段21とリンク解析手段22を用いて、ページ情報管理手段11に格納されているページの中から情報セットの代表ページとなるページを判定する。

【0082】ページ集合ソート手段23aは、図6に示すように、ページ集合取得手段23-1と、データ格納領域23-2と、情報セット取得手段23-3と、代表ページ取得手段23-4と、ソート手段23-5を備えており、入力装置40を介して渡されたサーチエンジン60によって検索された検索結果であるページ集合を、ページ情報管理手段11に格納されたページ重要度と情報セット管理手段13に格納されている情報セットデータを用いてソートすることにより、出力装置50に出力する。

【0083】次に、図7から図11を参照して本実施の形態の動作を図面を参照して詳細に説明する。

【0084】データ取込装置30がページ情報管理手段11とリンク情報管理手段12にデータ（ページ情報とリンク情報）を登録する処理と、ページ集合ソート手段23が入力装置40から入力されたURL集合をこれに対応する各URLのページ重要度にしたがってソートしこれを出力装置50であるディスプレイやファイルに出力する処理については、第1の実施の形態と同様である。

【0085】図7は、第2の実施の形態の重要度判定の処理の概要を説明するためのフローチャートである。

【0086】図7を参照すると、まず、ページ解析手段21がページ情報管理手段11に格納されているページを保持しているサーバのリストを作成し、これを代表ページ判定手段25のサーバリスト保持手段25-1に登録する（ステップ701）。

【0087】次に、代表ページ判定手段25のメンバURL取得手段25-2は、サーバリスト保持手段25-1からサーバを1つ取り出し、該サーバ上のURLリストをページ情報管理手段11から取得し、これを一時作業領域25-3に登録する（ステップ702）。

【0088】そして、一時作業領域25-3に格納されているサーバ上の各URLに対して、リンク解析手段22を用いて当該サーバ以外からのリンク参照数をカウン

トする（ステップ703）。

【0089】代表ページ判定手段25の代表ページ選択手段25-5は、サーバのルートディレクトリをデフォルトルートと設定し（ステップ704）、代表ページ発見の処理を実行する（ステップ705）。例えば、“http://www.nec.co.jp:80/DIRa/DIRb/index.html”のURLの場合、“www.nec.co.jp”がホスト、“www.nec.co.jp:80”がサーバ、その次の“/”がルートディレクトリ、“DIRa”がルート直下のサブディレクトリ、“DIRb”がDIRa直下のサブディレクトリを示す。また、index.htmlはページの実体をなすファイル名である。上記の場合、“/”がデフォルトルートとして設定される。

【0090】続いて、サーバ内のURLを再度取得し（ステップ706）、バウンダリを設定する（ステップ707）。ここで、バウンダリとは、上記のように、代表ページとそのメンバのページから構成されるある情報セットに関し、メンバとして適合するための条件を示す。

【0091】以上のステップ702からステップ707までの処理を、すべてのサーバに対し実行する（ステップ708）。

【0092】続いて、代表ページ発見の処理（ステップ705）について詳細に説明する。図8は、第2の実施の形態の代表ページ発見の処理を説明するためのフローチャートである。

【0093】図8を参照すると、代表ページ選択手段25-6は、ディレクトリ切出し手段25-5を用いて、与えられたデフォルトルート以下のURLを一時作業領域25-3から取得する（ステップ801）。ここでURLが存在しない場合はステップ811に移る（ステップ802）。

【0094】ここでURLが存在して取得した場合には、代表ページ選択手段25-6はこの取得したURLに対し代表ページ候補の選出処理を行う（ステップ803）。

【0095】代表ページ候補の選出（ステップ803）において、代表ページ候補が見つかった場合（ステップ804）は、これを代表ページ登録手段25-9を通じて、代表ページとして情報セット管理手段13に登録する（ステップ807）。

【0096】代表ページ候補が見つからなかった場合、ここでデフォルトルートがサーバのルートディレクトリである場合（ステップ805）には、ステップ801で取得したページをソートして、そのトップページを代表ページ登録手段25-9に登録する（ステップ806）。ステップ805でデフォルトルートがサーバのルートディレクトリでない場合には、ステップ811に移る。

【0097】さらに、代表ページ選択手段25-6は、ステップ807にて代表ページとして登録したページのサーバ外からの被リンク数をデフォルトルート以下のすべてのページのサーバ外からの被リンク数と比較し、代表ページの被リンク数が、デフォルトルート以下のページの被リンク数のうちの最大値である場合にはステップ809に、そうでない場合にはステップ811に移る(ステップ808)。

【0098】次に、ステップ808で代表ページの被リンク数が、デフォルトルート以下のページに対する被リンク数で最大値である場合には、代表ページ判定手段25-6は、ディレクトリ切出し手段25-5を用いて、デフォルトルート直下のサブディレクトリを切出し、さらに個々のサブディレクトリ内でサーバ外からの被リンク数が最大のページを求め、そのリンク数の値を合計して合計値を算出する(ステップ809)。ただし、各サブディレクトリにおける最大の被リンク数が規定値よりも小さい場合は、これを加算しない。

【0099】ここで算出した合計値と、代表ページの被リンク数を比較し(ステップ810)、合計値の方が小さい場合はステップ811に移り、そうでない場合には代表ページ発見処理を終了する。

【0100】次に、代表ページ選択手段25-6は、ディレクトリ切出し手段25-5を用いて、デフォルトルート直下のサブディレクトリの存在を調べる(ステップ811)。サブディレクトリが存在しないか、あるいはルートディレクトリを基準としてサブディレクトリの深さが設定値を超えた場合は、ホームページ発見処理を終了する。そうでなければ、デフォルトルート直下のディレクトリDを一時作業領域25-3から選択し(ステップ812)、代表ページ判定手段25-6は、ディレクトリ切出し手段25-5からサブディレクトリを順次取出し(ステップ813)、このサブディレクトリをデフォルトルートに再設定して(ステップ814)、代表ページ発見処理を再帰的に実行する(ステップ815)。サブディレクトリを全て処理するまでステップ813～ステップ815を繰返し(ステップ816)、完了したら代表ページ判定処理を終了する。

【0101】図9は、第2の実施の形態の代表ページ候補の選出の処理(ステップ803)を説明するためのフローチャートである。

【0102】図9を参照すると、代表ページ候補の選出の処理においては、代表ページ選択手段25-6が、ディレクトリ切出し手段25-5より与えられたURL集合に対して、サーバ外からの被リンク数が最大のページが存在するかを確認し(ステップ901)、存在すればそのページを代表ページ候補と設定する(ステップ904)。ステップ901で被リンク数最大のページが複数発見された場合は、これら複数の被リンク数最大のページに対して、サーバ内被リンク数カウント手段25-7

を用いてサーバ内からのリンク参照数をカウントしてその数を評価し(ステップ902)、サーバ外からの被リンク数が最大かつ、サーバ内からの被リンク数が最大であるページを代表ページ候補として設定する(ステップ905)。ステップ902の条件に合致するページが複数発見された場合は、これらのページのファイル名を、代表候補名格納手段25-8に格納されているファイル名と比較し(ステップ903)、ファイル名が合致した時点で、このページを代表ページ候補と設定する(ステップ906)。

【0103】次いで、バウンダリ判定の処理について図面を参照し詳細に説明する。図10は、第2の実施の形態のバウンダリ判定の処理を説明するためのフローチャートである。

【0104】図10を参照すると、バウンダリ判定の処理においては、まず、情報セット判定手段24のURLリスト作成手段24-1は、ページ情報管理手段11からサーバ内URLリストを取出し一時作業領域24-2に格納する(ステップ1001)。代表ページ選択手段24-3は、一時作業領域24-2に格納されているURLの中から代表ページと設定されているページのURLを選択し(ステップ1002)、この代表ページをディレクトリの深い順に取出す(ステップ1003)。

【0105】バウンダリ判定手段24-4は、代表ページが属するディレクトリを切出し、これを情報セット作成手段24-7にバウンダリとして登録する(ステップ1004)。親子探索手段24-5は、代表ページを起点とし、バウンダリとして設定されたディレクトリ内に存在するページを対象に幅優先探索を行い(ステップ1005)、検索によって得られたURLリストを情報セットのメンバ(ページ)として情報セット作成手段24-7に登録し、これらを一時作業領域24-2から取り除く(ステップ1006)。情報セット作成手段24-7は、上記の処理にて登録されたバウンダリとメンバを、情報セット管理手段13に設定する(ステップ1007)。

【0106】サーバ内のすべての代表ページについて、バウンダリとメンバの発見処理(ステップ1003～ステップ1007)を繰返す(ステップ1008)。

【0107】サーバ内のすべての代表ページについて、バウンダリとメンバの発見処理を実行した後、代表ページ選択手段24-3は、一時作業領域24-2にURLが残っているか調べる(ステップ1009)。残っていない場合は、本バウンダリ判定処理を終了する。残っている場合はURLを順次取出し(ステップ1010)、親ページを持つかどうかを調べる(ステップ1011)。

【0108】この一時作業領域24-2から取出したURLが親ページを持つ場合は、他セット移籍手段24-6は、親ページの属する情報セット(未定義の場合はサーバ)のメンバとして当該URLを情報セット管理手段

13に登録する(ステップ1012)。URLが親ページを持たない場合は、他セット移籍手段24-6は、該URLが属するサーバにおいてそのルートディレクトリをバウンダリとする情報セットのメンバとして、該URLに登録する(ステップ1013)。

【0109】以上の、ステップ1010～ステップ1013の処理を、一時作業領域24-2に残っている全てのURLに対し処理を行った後、本バウンダリ判定処理を終了する。

【0110】ページ集合ソートの処理について図面を参照し詳細に説明する。図11は、第2の実施の形態のページ集合ソートの処理を説明するためのフローチャートである。

【0111】図11を参照すると、ページ集合ソートの処理においては、まず、ページ集合ソート手段23のページ集合取得手段23-1が、入力装置40から入力されたURLリストを取得し、データ格納領域23-2に格納する(ステップ1101)。

【0112】次に、情報セット取得手段23-3は、データ格納領域23-2からURLを順次取り出し(ステップ1102)、情報セット管理手段13から該URLが属する情報セットを得て(ステップ1103)、該URLを該情報セットのメンバとしてデータ格納領域23-2に登録する(ステップ1104)。

【0113】ここで、該ページの持つページ重要度と、該情報セットの持つ情報セット重要度を比較し(ステップ1105)、ページ重要度が情報セット重要度よりも大きい場合には、このページ重要度を情報セット重要度に設定する(ステップ1106)。ここで、情報セットの重要度は、情報セットに属するページの重要度に基づいて種々の方法で設定することができる。例えば、情報セットに属するページの重要度のうち、最も大きな重要度をその情報セットの重要度として設定する。また、情報セットに属する各ページの重要度の合計をその情報セットの重要度として設定することも可能である。

【0114】上記ステップ1102からステップ1106の処理を、データ格納領域23-2中のすべてのURLに対し実行する(ステップ1107)。

【0115】次に、ソート手段23-5はデータ格納領域23-2に格納されている情報セットをその情報セット重要度でソートを実行し(ステップ1108)、この情報セット重要度の順に情報セットを取出す(ステップ1109)。

【0116】代表ページ取得手段23-4は、情報セット管理手段13から該情報セットの代表ページを取出し出力装置50に出力し(ステップ1110)、続いてソート手段23-5は、情報セット内ページをソート順に出力装置50に出力する(ステップ1111)。

【0117】ステップ1109からステップ1111を繰返し、データ格納領域23-2中のすべての情報セッ

トを出力装置50に出力する(ステップ1112)。

【0118】第1の実施の形態では、ステップ206においてリンク元URLとリンク先URLの近接性を同一サーバ上に存在しているかどうかで判定していた。また、入力装置から与えられたページ集合のみをソート出力の対象としていた。しかし、第2の実施の形態では、情報セット判定手段24が情報セットの範囲を、サーバだけでなくディレクトリやリンク関係も用いて判定して情報セット管理手段13に格納している。さらに代表ページ判定手段25は、各情報セットの代表ページを判定してする。また、ページ集合ソート手段23aは、情報セット単位でソートを行って出力するだけでなく、各情報セットの代表ページを付与する。

【0119】次に、本実施の形態の効果を説明する。

【0120】本実施の形態では、各ページのネットワークアドレスとリンク関係を用いて情報セットの範囲を判定する手段、判定した情報セットを記憶管理する手段、情報セットの代表ページを判定する手段、と情報セットを考慮してページ集合をソートする手段を備えるために、単にサーバの違いを考慮して得られるページ重要度よりも、情報セットを考慮する分だけ精度の高いページ重要度を得ることができ、情報セットを用いてソート結果をクラスタリングすることによってページ集合のソート結果の視認性を向上させ、さらに各ページの属する情報セットの特徴が容易に把握でき、さらに目的の情報セットのトップページを即座にアクセスすることができる効果を有する。

【0121】次に、本発明の第3の実施の形態について図面を参照して詳細に説明する。図12は、第3の実施の形態による重要度判定システム1の構成を示すブロック図である。

【0122】図12を参照すると、第3の実施の形態による重要度判定システムは、データ処理装置20が図3に示された第2の実施の形態におけるデータ処理装置20の構成に加え、情報木判定手段27を有する点、と第2の実施の形態におけるページ集合ソート手段23aを拡張したページ集合ソート手段23bを有する点で異なる。

【0123】図13は、本発明の第3の実施の形態の情報木判定手段27の構成を示すブロック図である。

【0124】情報木判定手段27は、ページ解析手段21とリンク解析手段22を用いて、情報セット管理手段13に格納されているすべての情報セットを対象とする情報木構造の判定を行う。

【0125】図13を参照すると、情報木判定手段27は、メンバ取得手段27-1と、一時作業領域27-2と、リンクカウント手段27-3と、親ページ確認手段27-4と、リンク距離カウント手段27-5と、親ページ登録手段27-6と、ディレクトリ取得手段27-7と、親度数カウント手段27-8と、ラベル付け手段

27-9を備える。

【0126】図14は、第3の実施の形態のページ集合ソート手段23bの構成を示すブロック図である。

【0127】ページ集合ソート手段23bは、入力装置40を介して渡されたページ集合を、ページ情報管理手段11に格納されたページ重要度、情報セット管理手段13に格納されている情報セットのデータを用いてソートし、出力装置50に出力する。

【0128】図14を参照すると、ページ集合ソート手段23bは、ページ集合取得手段23-1と、データ格納領域23-2と、ページ情報取得手段23-3aと、構造ソート手段23-5aと、親ページ確認手段23-6と、不要ページ削除手段23-7を備える。

【0129】次に、図15から図22を参照して本実施の形態の動作を図面を参照して詳細に説明する。

【0130】データ取込装置30がページ情報管理手段11とリンク情報管理手段12にデータを登録する処理については第1の実施の形態と、代表ページ判定手段25がページ解析手段21とリンク解析手段22を用いて各サーバの代表ページを判定する処理や、情報セット判定手段24がページ解析手段21とリンク解析手段22を用いて情報セットを判定して情報セット管理手段13に登録する処理については、第2の実施の形態と同様である。

【0131】代表ページと情報セットの判定後に行われる、各情報セットの情報木の判定の処理について図面を参照して詳細に説明する。図15は、情報木判定手段27による情報木の判定の処理を説明するためのフローチャートである。

【0132】図15を参照すると、まず、情報木判定手段27のメンバ取得手段27-1は、情報セット管理手段13から情報セットのリストを作成し（ステップ1501）、該リストから順次情報セットを順次取得し、さらに該情報セットのメンバURLを情報セット管理手段13から得て、これを一時作業領域27-2に格納する（ステップ1502）。

【0133】リンク距離カウント手段27-3は、一時作業領域27-2に格納されている情報セットのメンバURLの全てに対し、リンク解析手段22を用いてメンバURL間の被リンク数をカウントする（ステップ1503）。

【0134】親ページ確認手段27-4は、一時作業領域27-2に格納されているメンバURLの全てに対し、ページ管理手段11を用いてフレームページであるかどうかを確認し、URLがフレームページである場合、リンク管理手段12を用いて当該フレームページが内包するページ（子ページ）のURLを得る。親ページ登録手段27-6は、これら子ページURLの親として親ページURLを設定し、これをページ管理手段11に登録する（ステップ1504）。

【0135】そして、親ページ確認手段27-4は、ステップ1503にて設定したメンバURL間の被リンク数が”1”であるURLを選択し、親ページ登録手段27-6は、このリンク元ページのURLを親として設定し、これをページ管理手段11に登録する（ステップ1505）。

【0136】次に、リンク距離カウント手段27-5は、代表ページを起点として各メンバURLの親子関係を幅優先探索で調べ、代表ページからのリンク距離を計算し、リンク距離を一時作業領域27-2に設定する（ステップ1506）。

【0137】ディレクトリ取得手段27-7は、一時作業領域27-2に格納されているメンバURLが属しているサーバ内のディレクトリを取り出し、これをディレクトリの深い順にソートする。ただし、予め指定された深さ以上のものはディレクトリを取り出さないこととする（ステップ1507）。

【0138】そして、ディレクトリ取得手段27-7は、得られたディレクトリをその深さ順に取り出し、親度数カウント手段27-8に通知する（ステップ1508）。

【0139】次に、親度数カウント手段27-8は、リンク管理手段12を用いて、該ディレクトリに属しているメンバURLをリンクしているページ（親候補ページ）のURLを得て、各親候補ページから該ディレクトリに属するページへのリンク数を数えて、このリンク数を親候補ページの親度数とする（ステップ1509）。

【0140】親ページ確認手段27-4は親候補ページの中から親度数が最大のページを親ページとして選択し、親ページ登録手段27-6はの親ページからリンクされているメンバURLの親として親ページを設定し、親ページをページ情報管理手段11に登録する（ステップ1510）。

【0141】次に、親ページとして設定されたページの親度数をリセットし（ステップ1511）、親度数を持つ全てのページに対しステップ1510からステップ1511の処理を実行する（ステップ1512）。

【0142】さらに、ステップ1507で切出したすべてのディレクトリに対し、ステップ1508からステップ1512の処理を実行する（ステップ1513）。

【0143】次に、親度数カウント手段27-8は、ここで一旦すべてのメンバURLの親度数をリセットし、ステップ1513までの処理で得られたメンバURLの親子関係に基いて各メンバURLの親度数をカウントする（ステップ1514）。

【0144】ラベル付け手段27-9は、予め指定された閾値よりも大きな親度数を持つメンバURLに”INDEX”等の一意に識別できるラベルを付け、このラベルをページ情報管理手段12に登録する（ステップ1515）。

【0145】次に、ページ集合ソートの処理について図面を参照して詳細に説明する。図16は、第3の実施の形態のページ集合ソート手段23bによるページ集合ソートの処理を説明するためのフローチャートである。

【0146】図16を参照すると、ページ集合取得手段23-1は、入力装置40から入力されたURLリスト（ページ集合）を取得しデータ格納領域23-2に格納する（ステップ1601）。

【0147】続いて、情報セット整理の処理を実行する（ステップ1602）。

【0148】情報セット整理の処理が完了したら、親ページ確認手段23-6はデータ格納領域23-2において情報セットを指すポインタを順次更新し（ステップ1603）、ページ挿入の処理（ステップ1604）を実行する。

【0149】ページ挿入の処理が完了したら、不要ページ削除手段23-7はステップ1603において指定された情報セットを取出してページ削除の処理を実行する（ステップ1605）。

【0150】以上のステップ1603からステップ1605の処理を、全ての情報セットに対し実行する（ステップ1606）。

【0151】最後に、構造ソート手段23-5aは、構造ソートの処理を実行する（ステップ1607）。

【0152】情報セット整理の処理（ステップ1602）について図面を参照して詳細に説明する。図17は、第3の実施の形態による情報セット整理の処理を説明するためのフローチャートである。

【0153】図17を参照すると、第3の実施の形態による情報セット整理の処理は次のように実行される。すなわち、ページ情報取得手段23-3aは、データ格納領域23-2からURL（ページ）を順次取り出し（ステップ1701）、ページ情報管理手段11から該URLが属する情報セットとその親ページURLを得て（ステップ1702）、該URLを該情報セットのメンバとしてデータ格納領域23-2に登録する（ステップ1703）。

【0154】次に、該ページの持つページ重要度と該情報セットの持つ情報セット重要度を比較する（ステップ1704）。ページ重要度が情報セット重要度よりも大きい場合は、ページ重要度を情報セット重要度として設定する（ステップ1705）。

【0155】上記ステップ1701からステップ1705の処理を、データ格納領域23-2中のすべてのURLに対し実行する（ステップ1706）。

【0156】ページ挿入の処理（ステップ1604）について図面を参照して詳細に説明する。図18は、第3の実施の形態によるページ挿入の処理を説明するためのフローチャートである。

【0157】図18を参照すると、ページ挿入の処理

は、次のように実行される。すなわち、親ページ取得手段23-6はデータ格納領域23-2から、ステップ1603で指定された情報セットに含まれるメンバURLを順次取り出し（ステップ1801）、該メンバURLのラベルが“INDEX”であるかどうかを調べる（ステップ1802）。

【0158】ラベルが“INDEX”でない場合には、ステップ1801に戻り、次のメンバURLを順次取出す。

10 【0159】ラベルが“INDEX”である場合には、該メンバURLの親ページURLが該情報セットのメンバURLとしてデータ格納領域23-2に存在しているかを調べる（ステップ1803）。

【0160】親ページURLが存在している場合には、該メンバURLを該親ページURLの子として設定する（ステップ1804）。

【0161】親ページURLが存在しない場合には、ページ情報取得手段23-3aは該メンバURLの親ページの属する情報セットと、親ページURLと、ラベル値をページ情報管理手段11から取出してデータ格納領域23-2にメンバURLとして登録する（ステップ1805）。さらに登録した親ページURLの子として該メンバURLを設定する（ステップ1806）。

【0162】親ページ確認手段23-6は、上記ステップ1801からステップ1805を、該情報セットに含まれるすべてのメンバURLに対し実行する（ステップ1807）。

30 【0163】不要ページ削除の処理（ステップ1605）について図面を参照して詳細に説明する。図19は、第3の実施の形態の不要ページ削除の処理を説明するためのフローチャートである。

【0164】図19を参照すると、第3の実施の形態の不要ページ削除の処理は、次のように実行される。すなわち、不要ページ削除手段23-7はデータ格納領域23-2から、ステップ1603で指定された情報セットの代表ページを取出してこれを処理URLと設定し、さらに代表ページが属するディレクトリの深さdをd=0に設定する（ステップ1901）。

40 【0165】次に、非表示ページ選択の処理を実行する（ステップ1902）。ここで、非表示ページ選択の処理の流れにおいては、修理中に再び子ページに対して表示ページ選択の処理を実行する再帰的な処理を行っている。

【0166】非表示ページ選択の処理（ステップ1902）について図面を参照して詳細に説明する。図20は、非表示ページ選択の処理を説明するためのフローチャートである。

50 【0167】図20を参照すると、非表示ページ選択の処理は次のように実行される。不要ページ削除手段23-7は、該処理URLの子ページをデータ格納領域23

ー2から取り出し、これら子ページのディレクトリ深さ dc を $dc = d + 1$ と設定する (ステップ2001)。

【0168】該子ページと予め用意された非表示条件を比較し、条件に合致した子ページには非表示フラグを立てる (ステップ2002)。この非表示条件については、任意に設定されるものとする。

【0169】該子ページのうち、非表示フラグの立っていないものを選択し、ここから順次ページURLを取り出し (ステップ2003)、この子ページに対し非表示ページ選択処理を再帰的に実行する (ステップ2004)。

【0170】ステップ2003にて選択した全ての子ページに対し、ステップ2003とステップ2004の処理を実行する (ステップ2005)。

【0171】次いで、構造ソートの処理 (ステップ2005) について図面を参照して詳細に説明する。図21は、構造ソートの処理を説明するためのフローチャートである。

【0172】図21を参照すると、第3の実施の形態の構造ソートの処理は次のように実行される。構造ソート手段23-5aは、データ格納領域23-2に格納されている情報セットを情報セット重要度によりソートし (ステップ2101)、この情報セット重要度の順に情報セットを取出す (ステップ2102)。

【0173】次に、該情報セットの代表ページを選択して、これを処理ページと設定する (ステップ2103)。

【0174】そして、この処理ページURLを引数として出力処理を再帰的に実行する (ステップ2104～ステップ2106)。

【0175】出力処理 (ステップ2104) について図面を参照して詳細に説明する。図22は、本発明の第3の実施の形態の出力処理を説明するためのフローチャートである。

【0176】図22を参照すると、第3の実施の形態の出力処理は次のように実行される。まず、引数として渡された該処理ページURLの非表示フラグが立っているかを調べる (ステップ2201)。

【0177】非表示フラグが立っている場合には、この処理ページURLについては出力を行なうことなく出力処理を終了する。

【0178】非表示フラグが立っていない場合には、ページ情報を出力装置50に出力し (ステップ2202)、次に該処理ページURLに子ページが存在するかどうかを調べる (ステップ2203)。

【0179】子ページを持たない場合には、子ページの出力を行わないため出力処理を終了する。

【0180】子ページを持つ場合には、子ページURLを得て (ステップ2204)、これら子ページURLを順次取出し (ステップ2205)、出力処理を実行する

(ステップ2206)。

【0181】ステップ2204で得られた全ての子ページに対して、ステップ2205とステップ2206の出力処理を実行し (ステップ2207)、以上で出力処理を終了する。

【0182】第2の実施の形態では、情報セットの代表ページとバウンダリのみを判定しており、入力装置から与えられたページ集合を、情報セットの重要度と代表ページを考慮してソート処理を行っていた。これに対して本実施の形態では、情報木判定手段27が情報セットの構造をディレクトリやリンク関係も用いて判定してページ情報管理手段11に格納している。さらにページ集合ソート手段23bは、情報セット単位でソートを行い、さらに代表ページを付加して出力するだけでなく、各情報セットの構造を考慮して必要なページを付加し、不要なページを削除する。

【0183】次に、本実施の形態の効果を説明する。

【0184】本実施の形態では、各ページのネットワークアドレスとリンク関係を用いて情報セットの範囲を判定する手段、判定した情報セットを記憶管理する手段、情報セットの代表ページを判定する手段、情報セットないの情報木構造を判定する手段、と情報セットと情報木構造を考慮して、必要なページを追加し、不要なページを削除した上で、これらをソートする手段を備えるために、単にサーバの違いを考慮して得られるページ重要度よりも、情報セットを考慮する分だけ精度の高いページ重要度を得ることができ、情報セットを用いてソート結果をクラスタリングすることによってページ集合のソート結果の視認性を向上させ、さらに各ページの属する情報セットの特徴が容易に把握でき、目的の情報セットのトップページを即座にアクセスすることができ、さらにページ集合に含まれなくても必要なページは追加し、不要なページを削除することで重要なページだけを提示できる。

【0185】上述した情報セット重要度判定システムは、データ処理装置20、20a、20bの機能をハードウェア的に実現することは勿論として、コンピュータシステムの磁気ディスク、半導体メモリその他の記録媒体に記録された、上述した各機能を実現するための重要度判定用プログラムによってソフトウェア的に実現することも可能である。この重要度判定用プログラムは、図1、図3及び図12に示す記録媒体110からコンピュータシステムのデータ処理装置(CPU)に読み込まれ、データ処理装置の動作を制御することにより、上述した各手段の機能を実現して重要度判定の処理を実行する。

【0186】以上好ましい実施の形態及び実施例をあげて本発明を説明したが、本発明は必ずしも上記実施の形態及び実施例に限定されるものではなく、その技術的思想の範囲内において様々に変形して実施することができ

る。

【0187】

【発明の効果】第1の効果は、与えられたページリストを利用者にとって価値の高いものの順にソートすることができる。この結果、利用者が必要とするページ情報を上位に提示し、情報を選別する手間を削減することができる。

【0188】その理由は、同一テーマとみなせる個々のページを情報セットとしてグループ化し、この情報セットをまたがる参照関係をカウントするためである。

【0189】第2の効果は、与えられたページリストに適切な代表ページが含まれていない場合には、これを挿入してソートすることができる。この結果、検索時に適しなかった代表ページを利用者に提示することができる。

【0190】その理由は、情報セット外部からの参照数または情報セット内部からの参照状況に基いて各情報セットの代表ページを判定し、さらに与えられたURLリストを情報セット単位にグループ化してソートするためである。

【0191】第3の効果は、与えられたページリストから不要なページを削除できることにある。この結果、利用者に提示する検索結果の数を削減することができる。

【0192】その理由は、情報セット内の各ページ間の依存関係を判定して、提示するページとしなくても良いページを選別しているためである。

【図面の簡単な説明】

【図1】 本発明の第1の実施の形態による重要度判定システムの構成を示すブロック図である。

【図2】 本発明の第1の実施の形態の重要度判定の処理を説明するためのフローチャートである。

【図3】 本発明の第2の実施の形態による重要度判定システムの構成を示すブロック図である。

【図4】 本発明の第2の実施の形態の情報セット判定手段の構成を示すブロック図である。

【図5】 本発明の第2の実施の形態の代表ページ判定手段の構成を示すブロック図である。

【図6】 本発明の第2の実施の形態のページ集合ソート手段の構成を示すブロック図である。

【図7】 本発明の第2の実施の形態の重要度判定の処理を説明するためのフローチャートである。

【図8】 本発明の第2の実施の形態の代表ページ発見の処理を説明するためのフローチャートである。

【図9】 本発明の第2の実施の形態の代表ページ候補の選出の処理を説明するためのフローチャートである。

【図10】 本発明の第2の実施の形態のバウンダリ判定の処理を説明するためのフローチャートである。

【図11】 本発明の第2の実施の形態のページ集合ソートの処理を説明するためのフローチャートである。

【図12】 本発明の第3の実施の形態による重要度判

定システムの構成を示すブロック図である。

【図13】 本発明の第3の実施の形態の情報木判定手段の構成を示すブロック図である。

【図14】 本発明の第3の実施の形態のページ集合ソート手段の構成を示すブロック図である。

【図15】 本発明の第3の実施の形態の情報木の判定の処理を説明するためのフローチャートである。

【図16】 本発明の第3の実施の形態のページ集合ソートの処理を説明するためのフローチャートである。

【図17】 本発明の第3の実施の形態の情報セット整理の処理を説明するためのフローチャートである。

【図18】 本発明の第3の実施の形態のページ挿入の処理を説明するためのフローチャートである。

【図19】 本発明の第3の実施の形態の不要ページ削除の処理を説明するためのフローチャートである。

【図20】 本発明の第3の実施の形態の非表示ページ選択の処理を説明するためのフローチャートである。

【図21】 本発明の第3の実施の形態の構造ソートの処理を説明するためのフローチャートである。

【図22】 本発明の第3の実施の形態の出力処理を説明するためのフローチャートである。

【図23】 情報セットとバウンダリ概念を説明するための図である。

【符号の説明】

1 重要度判定システム

10 記憶装置

20 データ処理装置

30 データ取込装置

40 入力装置

50 出力装置

11 ページ情報管理手段

12 リンク情報管理手段

13 情報セット管理手段

21 ページ解析手段

22 リンク解析手段

23、23a、23b ページ集合ソート手段

23-1 ページ集合取得手段

23-2 データ格納領域

23-3 情報セット取得手段

23-3a ページ情報取得手段

23-4 代表ページ取得手段

23-5 ソート手段

23-5a 構造ソート手段

23-6 親ページ確認手段

23-7 不要ページ削除領域

24 情報セット判定手段

24-1 URLリスト作成手段

24-2 一時作業領域

24-3 代表ページ選択手段

24-4 バウンダリ判定手段

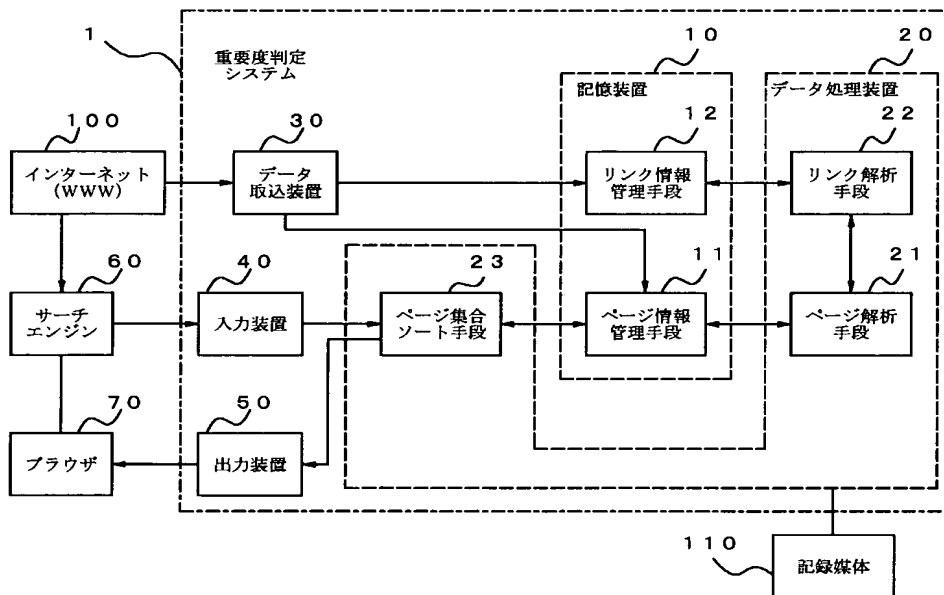
33

- 24-5 親子探索手段
- 24-6 他セット移籍手段
- 24-7 情報セット作成手段
- 25 代表ページ判定手段
- 25-1 サーバリスト保持手段
- 25-2 メンバURL取得手段
- 25-3 一時作業領域
- 25-4 サーバ外被リンク数カウント手段
- 25-5 ディレクトリ切出し手段
- 25-6 代表ページ選択手段
- 25-7 サーバ内被リンク数カウント手段
- 25-8 代表候補名格納手段

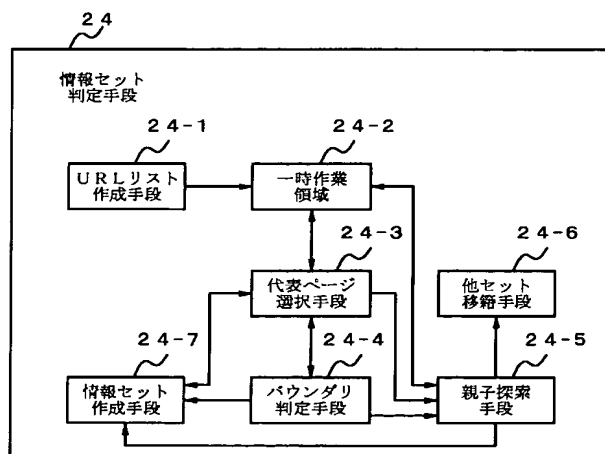
34

- 25-9 代表ページ登録手段
- 27 情報木判定手段
- 27-1 メンバ取得手段
- 27-2 一時作業領域
- 27-3 リンクカウント手段
- 27-4 親ページ確認手段
- 27-5 リンク距離カウント手段
- 27-6 親ページ登録手段
- 27-7 ディレクトリ取得手段
- 10 27-8 親度数カウント手段
- 27-9 ラベル付け手段
- 110 記録媒体

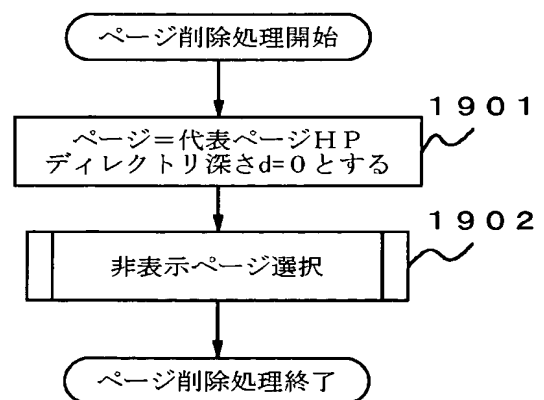
【図1】



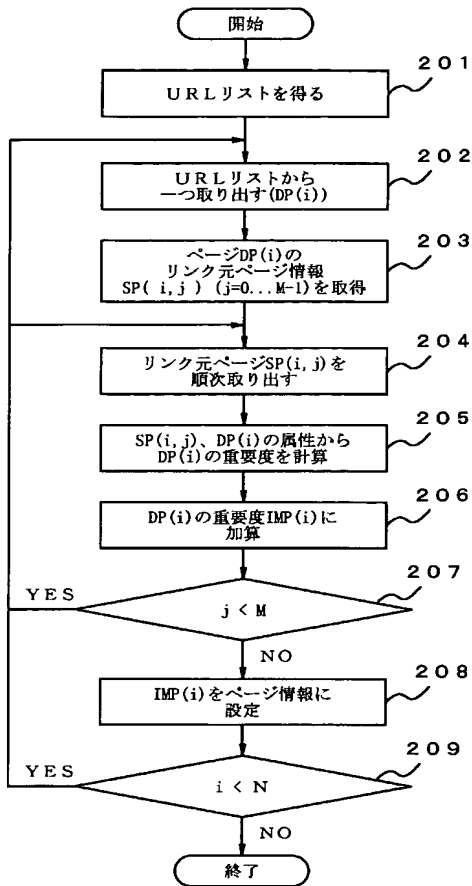
【図4】



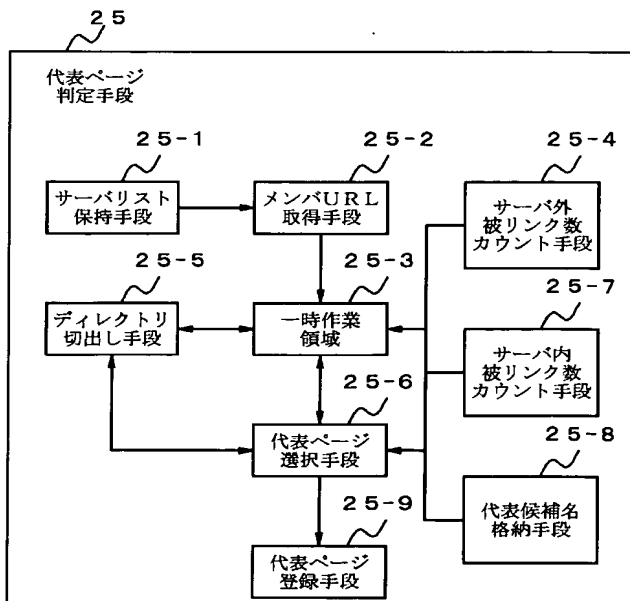
【図19】



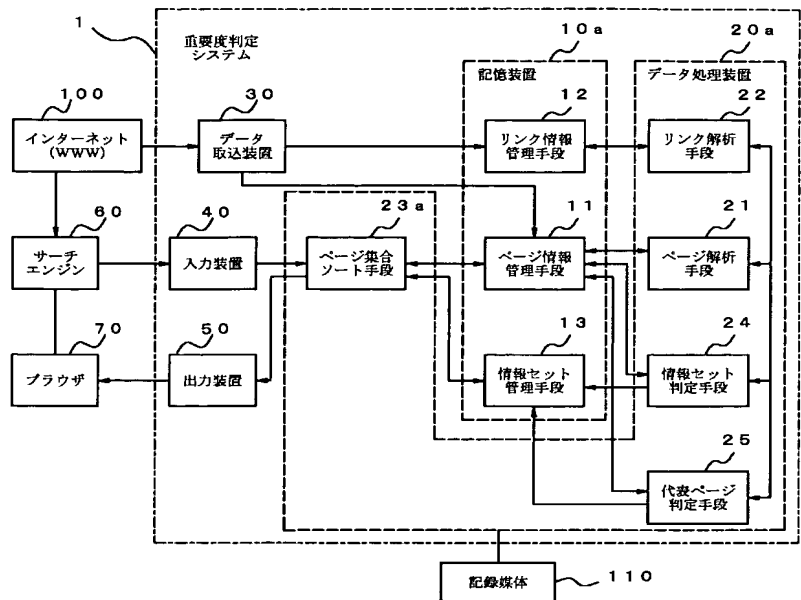
【図2】



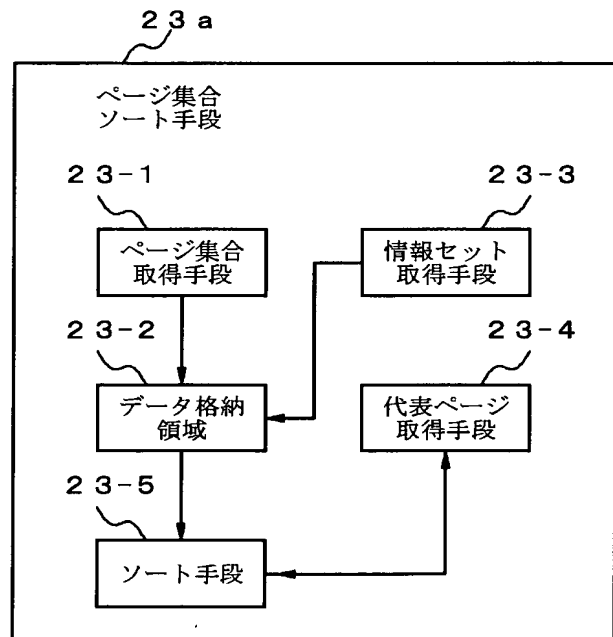
【図5】



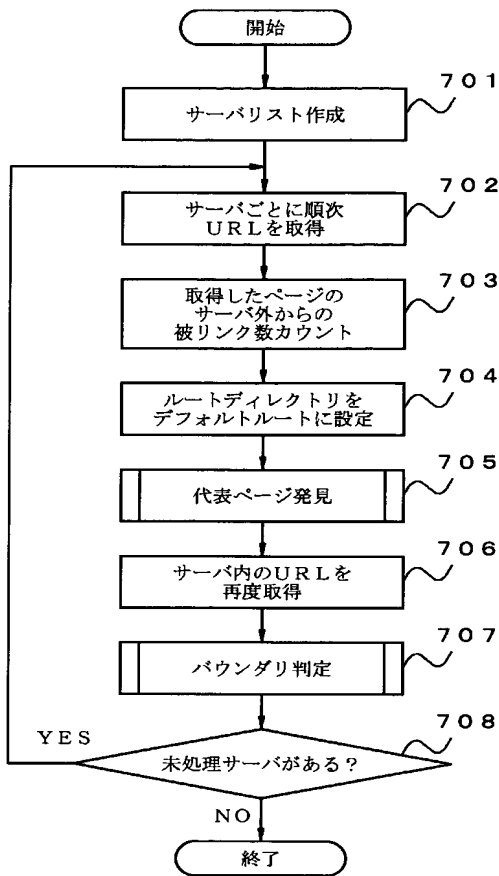
【図3】



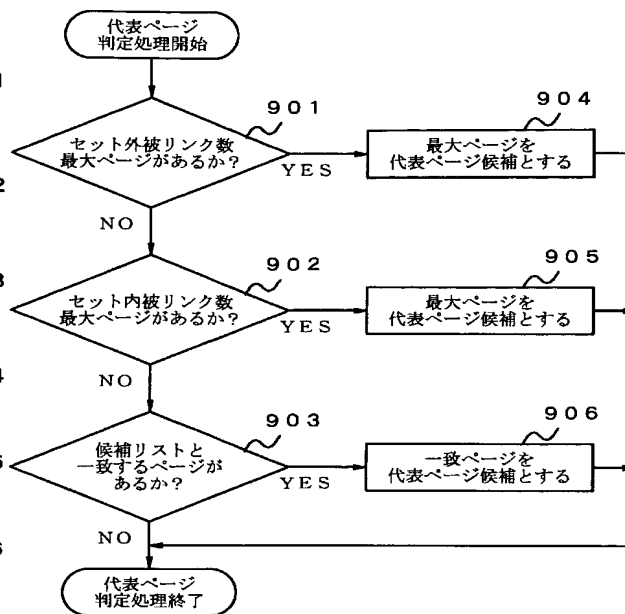
【図6】



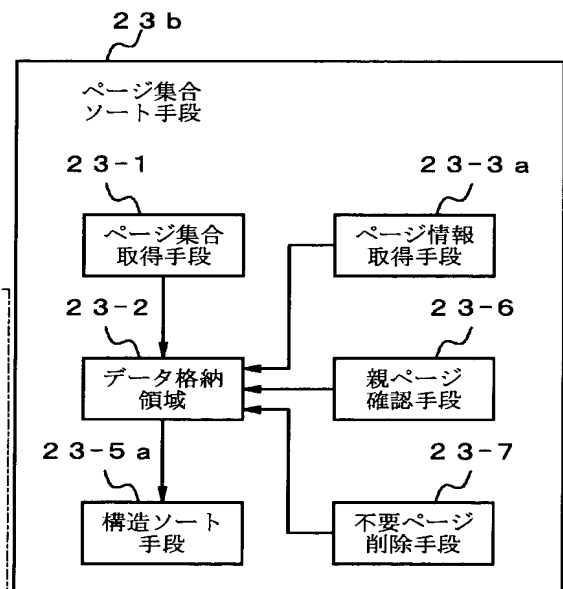
【図7】



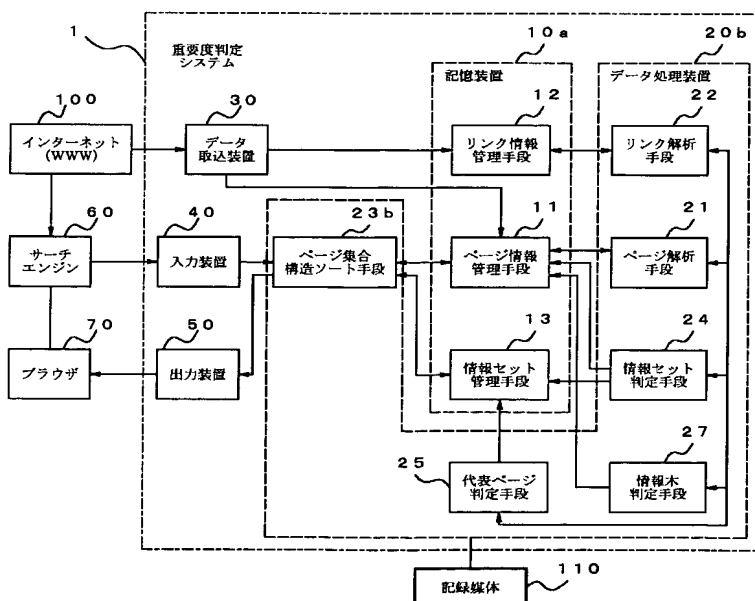
【図9】



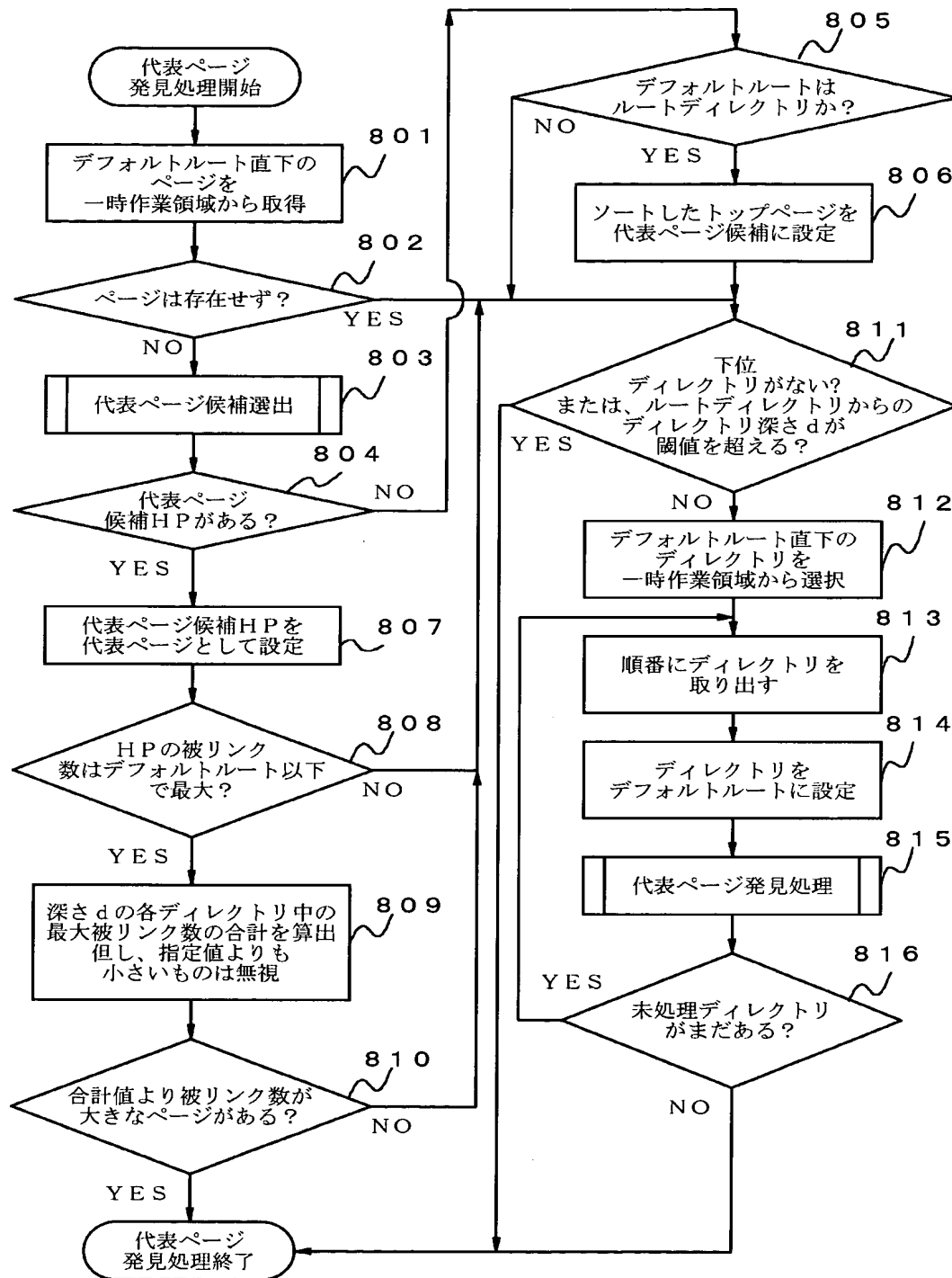
【図14】



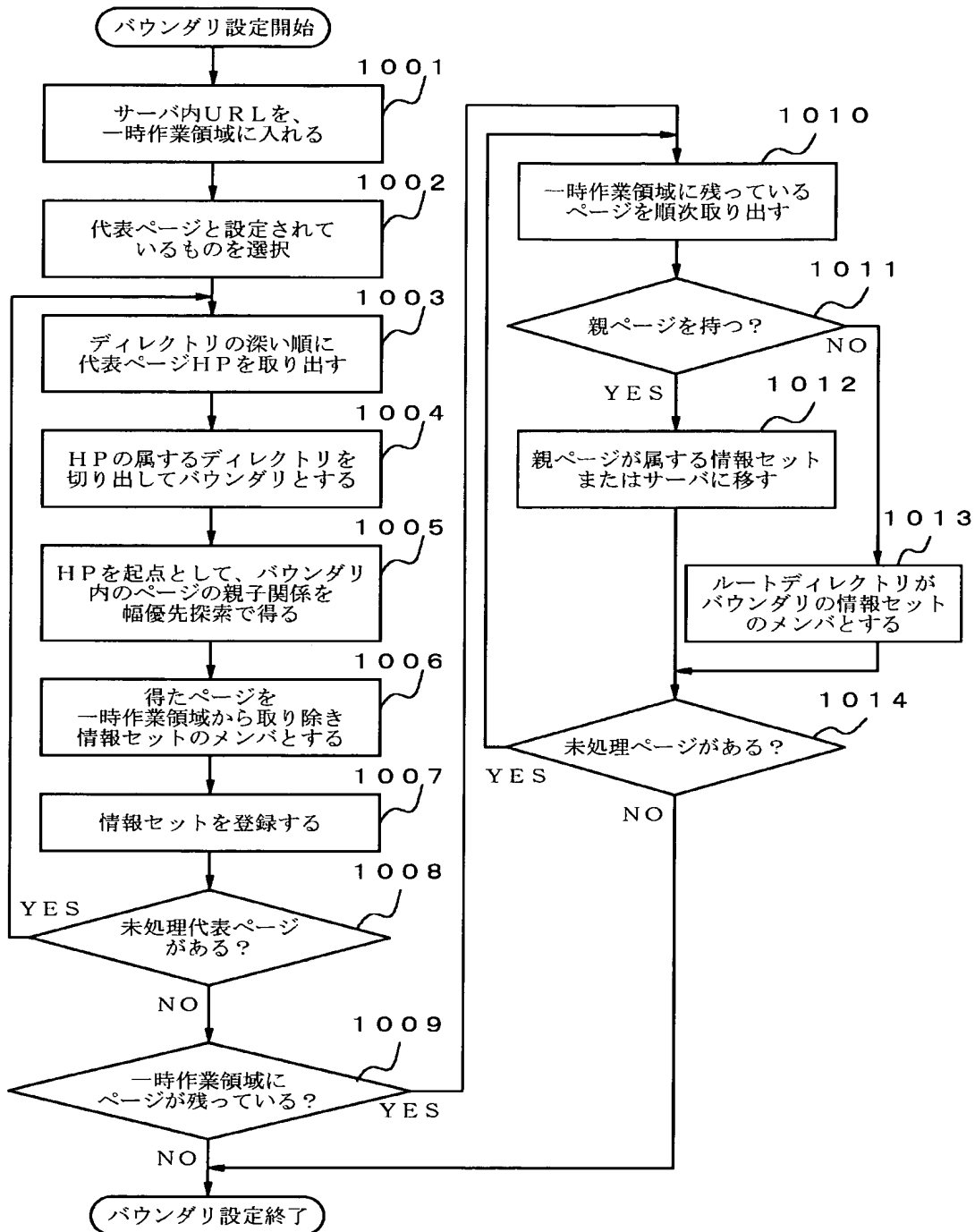
【図12】



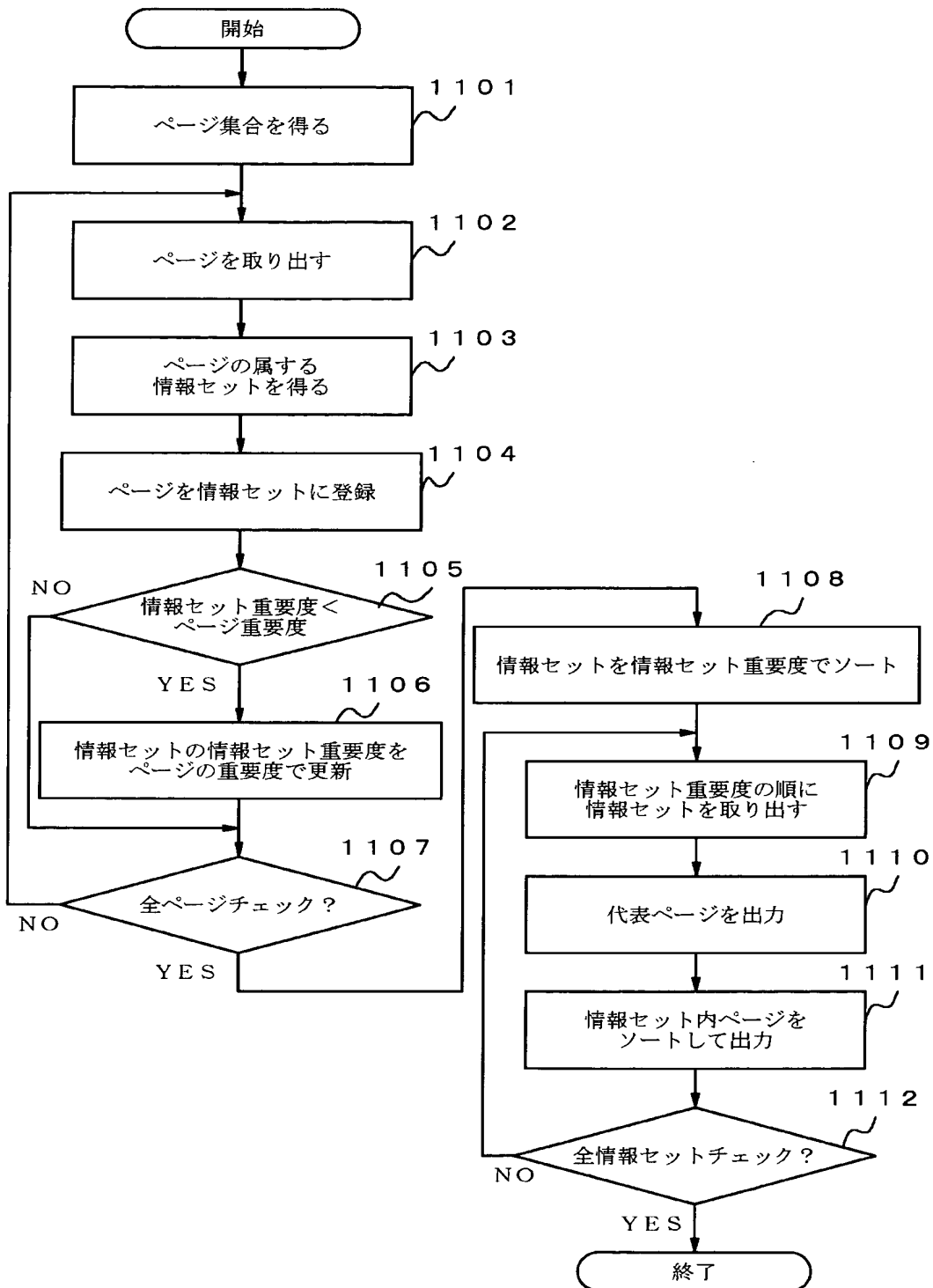
【図8】



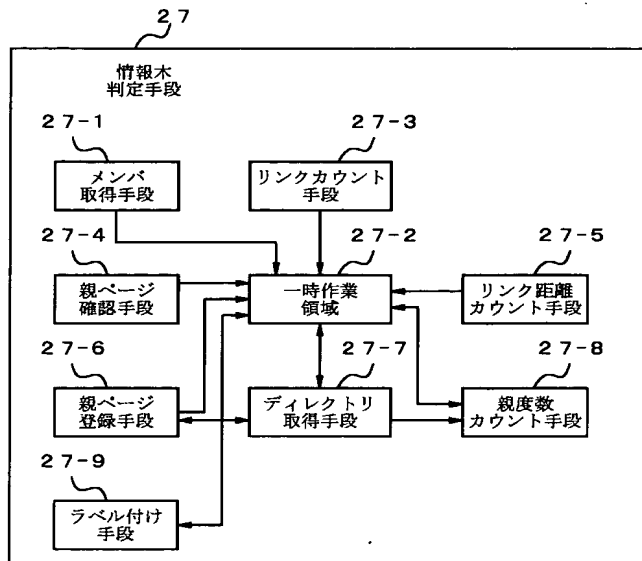
【図10】



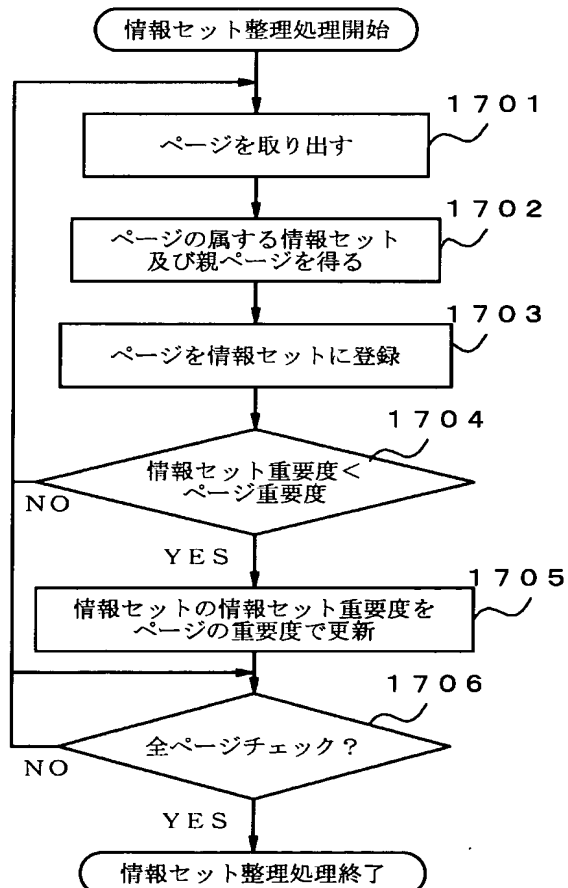
【図11】



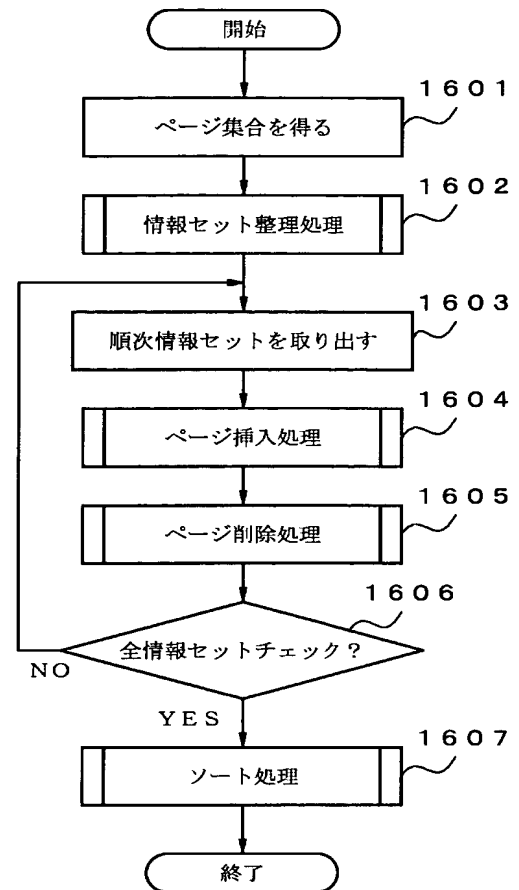
【図13】



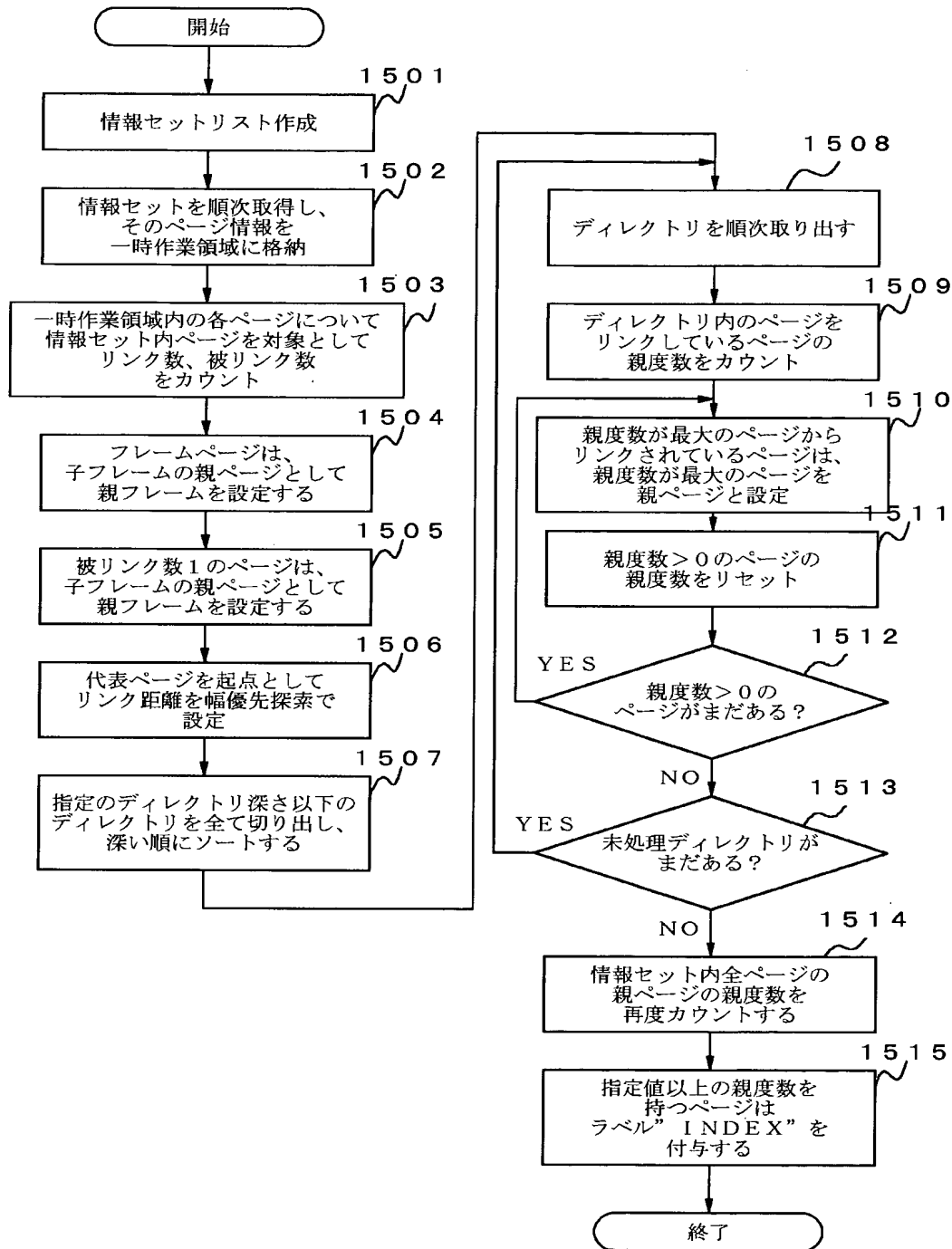
【図17】



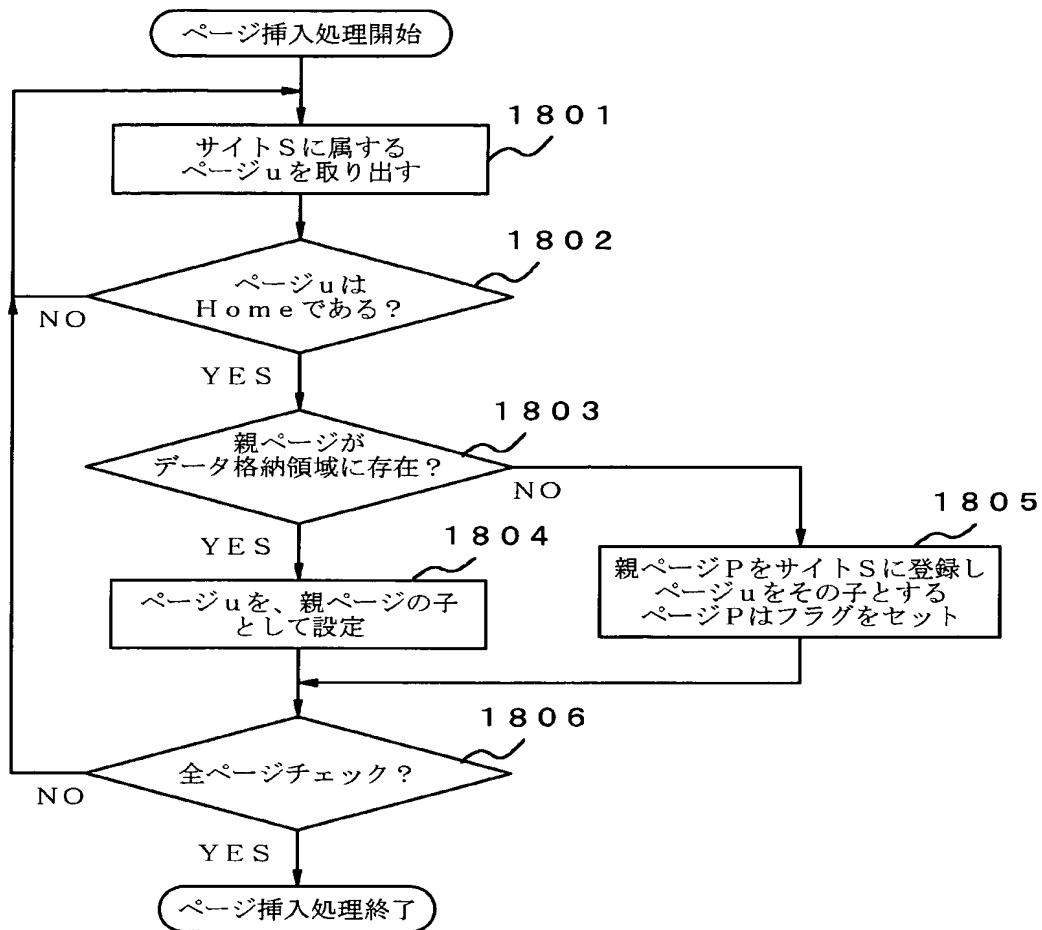
【図16】



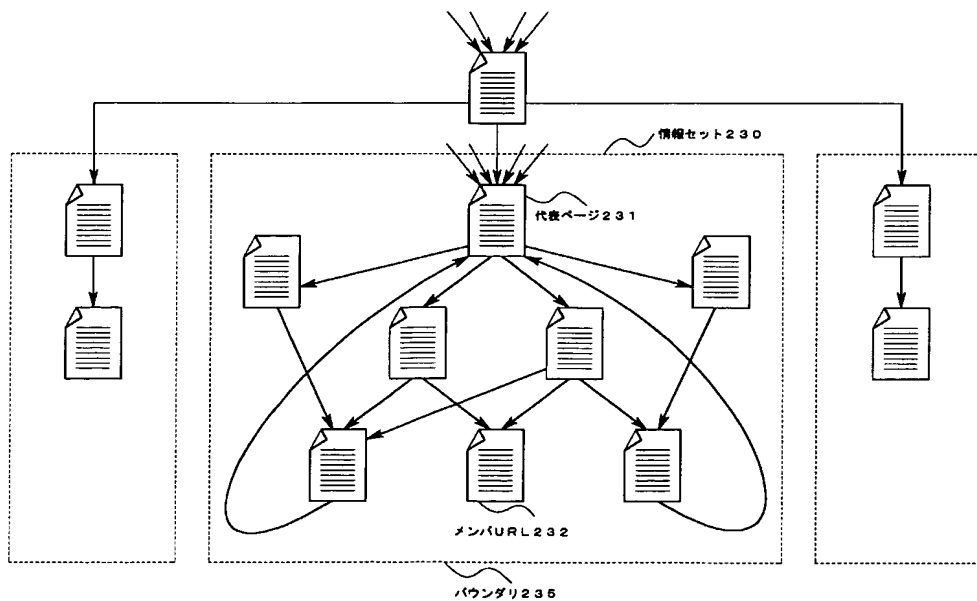
【図15】



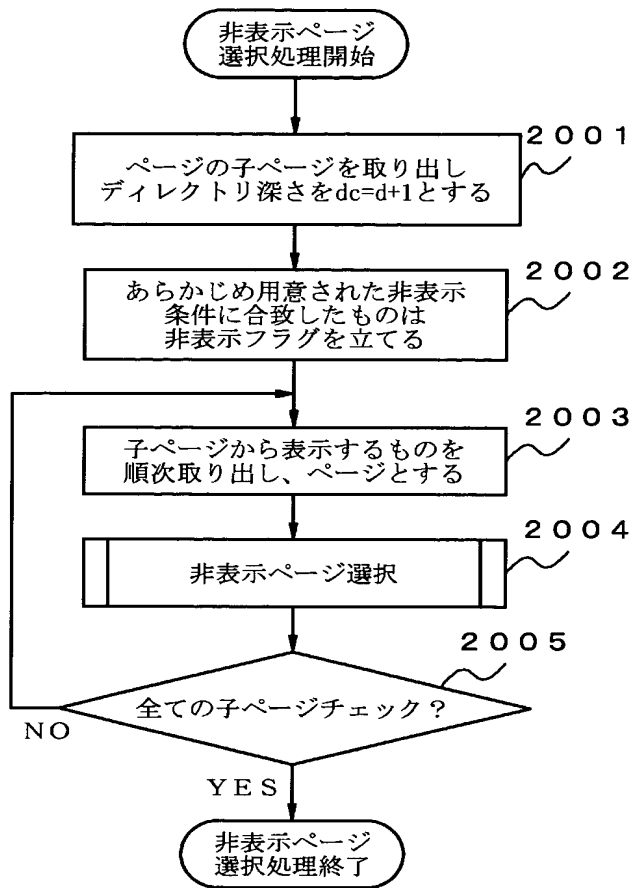
【図18】



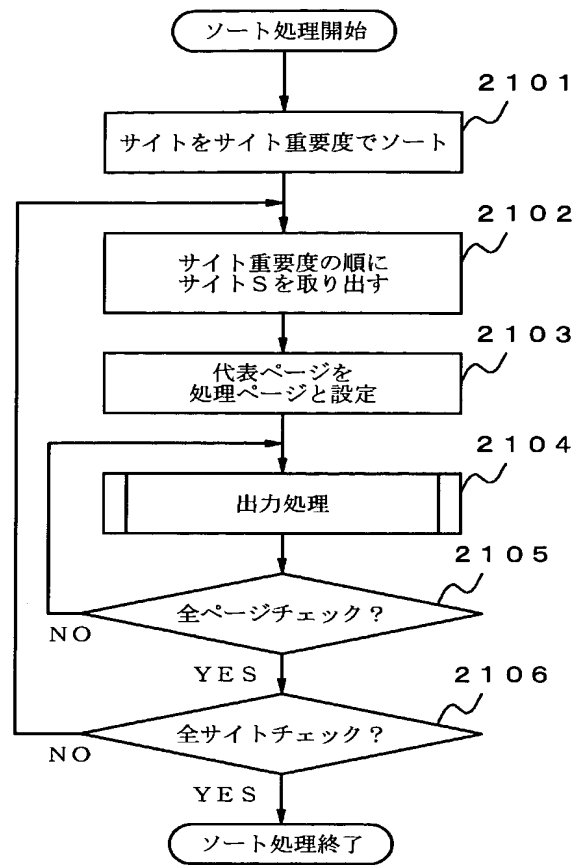
【図23】



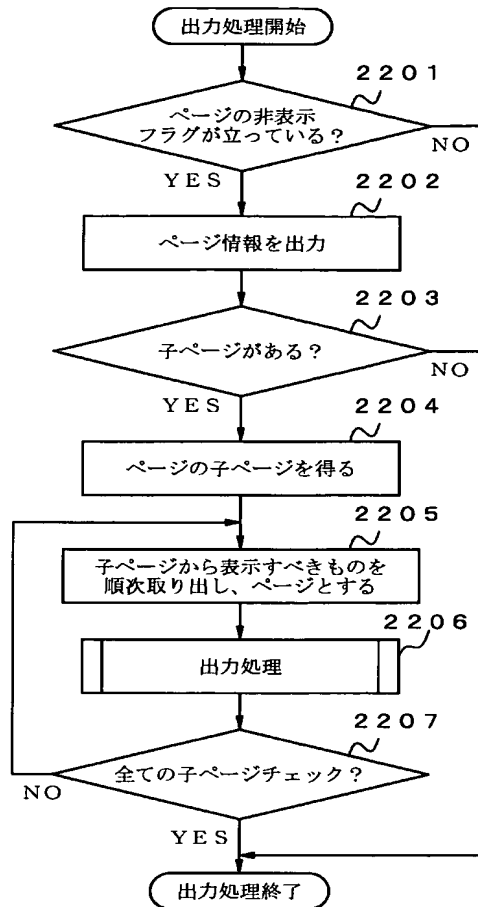
【図20】



【図21】



【図22】



【手続補正書】

【提出日】平成11年9月9日(1999.9.9)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】発明の名称

【補正方法】変更

【補正内容】

【発明の名称】 情報セット重要度判定システム及びその判定方法、及び情報セット重要度判定プログラムを記録した記録媒体